Designing an Open-World, Human-Level-Versatility Robot: A Safety-First Architecture for Mobility, Manipulation, and Social Competence

Douglas C. Youvan

doug@youvan.com

www.youvan.ai

October 24, 2025

Humanoid robotics is crossing a threshold: from choreographed demos toward real usefulness among people in unstructured environments. This paper proposes an end-to-end, safety-first specification for an open-world, human-level-versatility robot capable of operating in homes, clinics, schools, stores, and light-industry settings. We frame “human-level versatility” as a systems property that emerges only when mobility, dexterous two-hand manipulation, language-to-action grounding, tool use, and social compliance are jointly satisfied under strict safety and governance constraints. The contribution is a concrete, testable design: hybrid bipedal locomotion with wheel-assist, tendon-driven hands with tactile skins, unified metric-semantic world models, a diffusion plus task-and-motion planning stack for manipulation, and a policy engine that enforces lawful, ethical, and owner-scoped behavior. We define measurable acceptance tests, failure budgets, and audit receipts so that capability growth is paired with verifiable safety. A phased roadmap (P0 to P3) details how to mature from lab prototypes to public deployments, while a risk register highlights the hardest fronts: deformable manipulation, child and pet safety, knife and tool interlocks, tactile durability, and liability frameworks. The goal is to move beyond hype and sketches, providing a realistic blueprint that vendors, integrators, and regulators can collectively execute and evaluate.

Keywords: humanoid robotics, open world autonomy, dexterous manipulation, tactile sensing, hybrid locomotion, language grounding, task and motion planning, diffusion policies, safety kernel, speed and separation monitoring, tool interlocks, governance and audit, privacy by default, federated learning, acceptance tests, risk register, household robotics, service robotics, ethics and liability, standardization.

A collaboration with GPT-5. 83 pages. CC4.0.

Outline

1. Problem Statement and Mission Profile
   1.1 Domains of operation and design ceiling
   1.2 Non-goals and scope boundaries
2. System Requirements: From Use-Cases to Metrics
   2.1 Mobility, manipulation, language, social compliance
   2.2 Safety invariants, failure budgets, service levels
3. Physical Platform
   3.1 Form factor, mass, battery, ingress protection
   3.2 Actuators, series elasticity, backdrivability
   3.3 Tactile skin architecture and durability
4. Perception and World Modeling
   4.1 Multi-camera, depth, event, audio, and force sensing
   4.2 Metric-semantic maps, scene graphs, uncertainty fields
   4.3 Open-set detection and outlier handling
5. Planning and Control
   5.1 Locomotion: footstep plus wheel-assist planners
   5.2 Manipulation: diffusion policies with TAMP scaffolding
   5.3 Tactile servoing and non-prehensile skills
6. Language-to-Action and Human Interaction
   6.1 Instruction DSL, confirmations, and repair
   6.2 Proxemics, gaze, gesture, accessibility modes
7. Safety Kernel, Policy Engine, and Governance
   7.1 Speed and separation monitoring, force ceilings
   7.2 Tool interlocks, lawful and owner-scoped policies
   7.3 Audit receipts, black-box buffers, privacy by default
8. Learning Pipeline and Data Stewardship
   8.1 Foundation visuomotor models and robot demos
   8.2 On-device adaptation, federated learning, red-teaming
   8.3 Dataset provenance and skill store signing
9. Reliability, Maintainability, and Cost
   9.1 MTBF targets, hot-swap modules, consumables
   9.2 Service workflows and total cost of ownership

10. Risk Register and Mitigations
    10.1 Deformables, open-set safety, tool hazards
    10.2 Supply chain, floor damage, voice spoofing, liability

11. Validation Suite and Acceptance Tests
    11.1 Mobility trials, manipulation cycles, social safety
    11.2 Household “new home in a day” and “errand loop”

12. Deployment Roadmap (P0 to P3)
    12.1 Lab to pilot homes
    12.2 Back-of-house services
    12.3 Broad public deployments and certification

13. Ethics, Law, and Standards
    13.1 Consent, privacy, and data minimization
    13.2 Alignment with ISO/ANSI safety norms
    13.3 Insurance and incident response

14. Related Work and Differentiation
    14.1 Humanoid trends and gaps
    14.2 Why a safety-first, audit-ready stack is decisive

15. Conclusion: From Demos to Duty Cycles
    15.1 What success looks like
    15.2 Next experiments and community benchmarks

References

Appendices: Interface specs, DSL grammar, simulator assets, bill of materials, extended test protocols.

1. Problem Statement and Mission Profile

The goal is a generalist humanoid system that can execute multi-step tasks safely among people in **unstructured, everyday environments**. “Human-level versatility” is scoped operationally: the robot must (a) navigate real spaces with stairs, clutter, and moving actors; (b) perform two-hand manipulation over rigid and deformable objects; (c) ground natural-language instructions into reliable action sequences; (d) use common household and light-industrial tools; and (e) comply with social norms and hard safety constraints. The **mission profile** assumes variable lighting, partial occlusions, background noise, pets and children, and changing layouts. Success is measured not by stunt demos but by **duty-cycle reliability** (hours between interventions), **safety incident rate** (near-misses, force/energy exceedances), and **task completion under variation** (same intent, different geometry/objects).

Key operating assumptions:

- **Mixed autonomy:** on-device policies handle the common case; rare edge cases escalate to a human overseer.
- **Local-first privacy:** no hard dependency on cloud for safety-critical perception, planning, or actuation.
- **Incremental deployment:** begin with fenced/geo-fenced modes and trusted homes/sites, broaden as metrics clear thresholds.
- **Tool interoperability:** prefer adapters and software “plug-ins” over custom fixtures; the robot adapts to the world, not vice versa.

1.1 Domains of operation and design ceiling

Primary domains (intended from V*HLV-1.0 onward):

- **Homes and apartments:** kitchens, laundry rooms, living spaces, bedrooms; tasks include dish handling, laundry fold/sort, tidying, light food prep with guards, trash/recycling, making beds, basic cleaning and organizing.
- **Back-of-house services:** hotel housekeeping carts and storerooms, big-box restocking at night, clinic supply runs, office facilities support; tasks include tote/bag movement, shelf stocking, linen handling, light janitorial resets, inventory scans.
- **Community spaces and buildings:** lobbies, corridors, elevators, classrooms; tasks include deliveries, room resets, moving furniture on casters, AV setup, signage placement, event turnover.

- **Light industry and makerspaces:** guarded or semi-guarded cells near people; tasks include line-feeding, kitting, fixture operation with force/velocity limits, EoL inspection assists, rework prep.

Environmental ceiling (design targets):

- **Mobility:** 20 cm step, stairs to standard residential code, 12 deg ramps, wet tile, low-pile carpet, outdoor sidewalk for short errand loops (IP54 kit).
- **Dexterity:** one-hand dynamic payload 5 kg, two-hand cooperative lift 20 kg torso-assisted; reliable grasps on rigid objects 2–40 cm, deformables (cloth/cables) with tactile servoing.
- **Perception:** robust to 20–500 lux lighting, partial occlusions, reflective surfaces; open-set detection with conservative fallback.
- **Social compliance:** proxemics, yielding, gaze/gesture grounding, voice e-stop, signed task receipts; predictable behavior around children/pets.
- **Autonomy & uptime:** 3–6 h active runtime per pack, hot-swap in 20 min, mean fall rate < 1 per 200 h in mixed home tests, < 1 tele-assist per active hour in target domains after commissioning.

Task ceiling (illustrative, not exhaustive):

- **Household:** unload dishwasher→sort→put-away; laundry wash→dry→fold→drawer placement; daily tidy; prepare simple meals with cut shields; wipe counters; take out trash.
- **Service:** room turnover (strip/make beds, reset towels, restock amenities), off-hours shelf restock, cart loading/unloading, conference room reset.
- **Light industry:** pick parts from bins, stage sub-assemblies, operate guarded levers/buttons, present parts to camera/fixture, pack-out and labeling.

1.2 Non-goals and scope boundaries

To keep the system safe, buildable, and auditable, the following are **explicitly out of scope** for initial generations:

A. Medical and high-risk procedures

- **No invasive medical acts** (injections, suturing, surgery).

- **No unsupervised caregiving** that requires medical judgment (administering meds, lifting/transporting people, bathing).
- Rationale: unacceptable risk concentration; certification regime and liability do not support this at launch.

B. Hazardous environments beyond the platform's protection class

- **No firefighting**, CBRN response, or heavy construction/demolition.
- **No operation in explosive atmospheres** (ATEX) without a certified variant.
- Rationale: requires specialized chassis, sealing, sensors, and doctrine.

C. Open-ended childcare or pet care as primary function

- The robot **must remain a tool**, not a caregiver or babysitter; it may **yield and avoid**, never assume supervisory responsibility.
- Rationale: ethical and legal boundaries; social trust and certification absent.

D. Weapons, coercion, or security enforcement

- **No weapon handling**, restraint, or enforcement roles.
- **No face recognition for policing** or person-specific pursuit.
- Rationale: misuse risk, misidentification, unacceptable amplification of harm.

E. Fully unsupervised outdoor navigation in traffic

- **No street-level navigation among vehicles** as a baseline feature; constrained "errand loops" only (sidewalk→store counter→return) with strict routes and fallbacks.
- Rationale: urban traffic is a separate autonomy class; risk and regulation differ.

F. Unlimited improvisation or self-modification

- **No autonomous tool invention** that bypasses interlocks; **no self-modifying safety envelopes**; skills must be signed and provenance-tracked.
- Rationale: preserve auditability and predictable safety margins.

G. Continuous surveillance or data brokerage

- **No default video/audio recording** beyond transient buffers; **no data resale**; all learning is local-first with explicit opt-in for federated updates.
- Rationale: privacy by default; maintain social license to operate.

H. Fragile or high-value artisan tasks without fixtures

- **No museum-grade art handling**, precision lutherie, or watchmaking.
- Rationale: tactile resolution and liability not aligned with such tasks initially.

I. Unlimited cost reduction at the expense of safety

- **No disabling of force/velocity ceilings**, removal of e-stops, or policy guardrails to reach price points.
- Rationale: safety kernel is non-negotiable.

Boundary posture (what happens when the world exceeds scope):

- The robot must **refuse** tasks violating policy or safety envelopes, **de-energize** dangerous tools, and **escalate** with a clear, auditable explanation: what constraint was hit, what alternatives exist (e.g., request a fixture, slower mode, human assist), and how to proceed safely.

Implication for readers/practitioners: By stating domains, ceilings, and non-goals up front, we anchor design trade-offs (sensors, actuators, compute, safety kernel) and align stakeholders—engineers, integrators, regulators, insurers, and end-users—on what "human-level versatility" means **in practice**, not in myth.

2. System Requirements: From Use-Cases to Metrics

This section turns mission goals into verifiable requirements. Each capability is defined by: (a) use-case families, (b) measurable metrics with target and minimum-viable thresholds, (c) evaluation protocols, and (d) instrumentation needed to audit results. Numbers are initial stakes in the ground; they should tighten through pilots without eroding safety.

2.1 Mobility, manipulation, language, social compliance

2.1.1 Mobility (biped + wheel-assist)

Use-cases: home stairs and thresholds, wet tile, carpet, cluttered corridors, elevator entry/exit, curb cuts, short sidewalk segments.

Metrics and targets

- Gait stability: mean fall rate < 1 per 200 h mixed-home testing (target), < 1 per 100 h (minimum).

- Push-recovery: recover from 80 N lateral impulse at pelvis height in < 250 ms with no step loss in 95 percent of trials.
- Terrain handling: stairs to residential code up and down while carrying 5 kg in one hand with 0 incidents over 100 ascents and 100 descents.
- Ramp handling: 12 deg ramps, forward and backward, success 99 percent over 200 traversals.
- Micro-navigation: doorway and elevator ingress/egress first-try success 97 percent over 300 trials across 10 buildings.
- Energy budget: 300 to 600 W average while locomoting indoors at 0.8 m s, including perception and safety loops.

Evaluation protocol

- Mixed-home course with randomized obstacles; impulse tests with instrumented pendulum; 10-building elevator benchmark; continuous logging of base pose, foot forces, slip flags at 500 Hz.

2.1.2 Manipulation (two-hand dexterity)

Use-cases: dish handling, laundry fold and sort, cable routing, opening containers, shelf restock, tote kitting, lever and button actuation in guarded cells.

Metrics and targets

- Rigid-object grasp success: > 98 percent on objects 2 to 40 cm across, 1.5 average grasp attempts, over 1k-cycle runs.
- Deformables: laundry fold F1 >= 0.90 over 200 garments mixed types; cable threading success >= 95 percent with 3 m flexible cords; cloth wipe coverage >= 95 percent with streak defects < 3 percent of area.
- Force control: peak contact overshoot < 10 N beyond planned envelope on fragile tasks; slip detection latency < 15 ms via tactile skin.
- Payload: dynamic one-hand 5 kg, two-hand cooperative 20 kg torso-assisted; hold durations 60 s without regrasp.
- Tool use: quick-change wrist adapters identified in < 200 ms; interlock validation before torque enable 100 percent.

Evaluation protocol

- Kitchen suite (dishwasher load and unload, cutlery sorting, glassware handling); laundry suite (wash to fold to drawer); cable and zipper benchmarks; force-limited tests with instrumented dummies; repeatability runs of 100 cycles per task.

2.1.3 Language-to-action and task grounding

Use-cases: owner issues multi-step instructions, robot confirms ambiguities, executes, and reports outcomes with receipts.

Metrics and targets

- Instruction success: > 95 percent on a 200-instruction compositional set with object references, spatial relations, and constraints.
- Repair dialogs: successful recovery after a single correction in > 90 percent of cases.
- Latency: intent parse to first action start < 300 ms median; confirmation round-trip < 1.0 s median.
- Program synthesis: instruction to internal DSL translation correctness > 98 percent on held-out prompts.
- Provenance: 100 percent of actuator uses have signed receipts recording who, what, when, where, and force ranges.

Evaluation protocol

- Blind, randomized command lists across 10 homes; disfluency and accent robustness with 20 speakers; background-noise tracks at 45 to 70 dBA; offline DSL gold sets for grading.

2.1.4 Social compliance and human factors

Use-cases: operating near people, pets, and property; yielding, maintaining distance, attention signaling, and accessibility modes.

Metrics and targets

- Proxemics: maintain personal-space buffers according to context rules in 99 percent of encounters; auto-yield success > 99 percent.

- Child and pet safety: zero unsafe contacts over 1k h of mixed testing with mobile mannequins and robotic pet proxies; speed and separation monitoring engages preemptively in 100 percent of encroachment events.
- Accessibility: ASL subset command accuracy > 95 percent; large-print UI task confirmation error < 1 percent.
- Attention signaling: gaze and light cues acknowledged by participants as “clear” in > 95 percent of human studies.

Evaluation protocol

- Human-in-the-loop corridors with scripted and unscripted crossings; proxemics maze; pet-proxy dart tests; IRB-reviewed user studies for cue clarity.

2.2 Safety invariants, failure budgets, service levels

2.2.1 Safety invariants (must hold at all times)

1. **E-stop invariants**
    - Physical and voice e-stop halts motion to safe state in < 500 ms median, < 800 ms worst-case.
    - E-stop is latched and requires explicit human reset.
2. **Energy and force limits**
    - Tool power rails gated by interlocks; no energization without tool ID and policy permit.
    - Joint torque ceilings enforceable by kernel; contact force outside envelope triggers reflex deceleration in < 10 ms.
3. **Speed and separation monitoring (SSM)**
    - Dynamic speed caps scale with human proximity; violation probability < 1e-6 per hour under certified sensor health.
4. **Open-set detection and conservative fallback**
    - Unknown or low-confidence object classes drive speed reduction and task refusal; no uncontrolled novel-contact behaviors.

5. **Data minimization and privacy**
    - No persistent audio or video recording by default; rolling black-box buffer of 120 s maximum, encrypted and operator-held.
6. **Policy guardrails**
    - Illegal or unsafe commands are refused with rationale; no self-modification of safety envelopes; all skill loads signed and verified.
7. **Fail-safe on degraded sensing**
    - If critical sensors fail (skin, near-field depth, IMU), platform enters reduced-capability mode or safe-stop depending on context.

2.2.2 Failure budgets (explicit allowances and triggers)

We budget failures so that safety is never traded for throughput. Budgets are per 100 operating hours unless noted.

- **Falls:** budget 0.5 (target 0.2). Breach triggers mobility regression analysis and firmware freeze.
- **Unsafe contacts:** budget 0.0. Any breach halts program and invokes incident protocol.
- **Tele-assists:** budget <= 60 total minutes (target <= 30) across interventions; any single task with intervention rate > 10 percent is flagged for redesign.
- **Task aborts due to low confidence:** budget <= 3 percent; higher rates indicate perception or policy tuning needs.
- **False refusals of safe tasks:** budget <= 2 percent; trade off against false accepts is tuned toward safety.
- **Tool interlock faults:** budget 0; any tool energization without ID or policy is a critical event.
- **Sensor health downgrades:** budget <= 1 per 100 h; beyond this, maintenance cadence is insufficient.

2.2.3 Service levels (SLOs and SLIs)

Availability SLOs

- Core autonomy availability: 99.0 percent per 30-day window for home deployments, 99.5 percent for back-of-house deployments.
- Safety-kernel availability: 99.99 percent; any kernel fault forces safe-stop and is counted separately.

Performance SLIs

- Mean time between interventions (MTBI): >= 8 h in target domains after commissioning.
- Mean time to safe-stop on e-stop: < 0.5 s median.
- Mean task cycle success: >= 97 percent in acceptance test suites.
- Battery hot-swap downtime: <= 20 min 95th percentile.

Maintainability

- Mean time to repair (MTTR) for swappable modules: hand 5 min, arm 30 min, battery 5 min.
- Consumable replacement intervals: fingertip pads 100 h light-home use, soles 200 h, skin patches 300 h.

Telemetry and audit

- 100 percent actuator events logged with timestamp, pose, force bands, tool IDs, and policy version.
- Tamper-evident logs signed on-device with rotating keys; daily rollovers; owner-controlled export.

2.2.4 Test harness and acceptance criteria (go/no-go)

A deployment is approved only if all below pass on the intended site class.

- **Mobility suite:** mixed-course traversal 2 h continuous, 0 falls, 0 unsafe contacts, push-recovery pass at 80 N, elevator benchmark pass rate >= 97 percent.
- **Manipulation suite:** kitchen set (dishwasher, glassware) success >= 97 percent; laundry set F1 >= 0.90; cable tasks >= 95 percent; force overshoot violations 0.

- **Language suite:** 200-instruction set success >= 95 percent; correction recovery >= 90 percent; refusal accuracy for unsafe asks >= 99 percent.
- **Social suite:** proxemics compliance >= 99 percent; child and pet proxy tests 0 incidents; e-stop tests within latency targets.
- **Safety kernel:** red-team triggers for tool interlocks, SSM breaches, sensor downgrades all drive correct mitigations 100 percent of the time.
- **Privacy checks:** no data egress without explicit opt-in; black-box buffer encryption verified; audit receipts complete for all tasks.

2.2.5 Instrumentation requirements

To make metrics real, the system must ship with:

- High-rate joint torque and position logs (1 kHz control loop, 100 Hz archived).
- Tactile skin summaries (contact patches, centroid, slip flags) at 200 Hz.
- Safety-kernel tracepoints for speed and separation decisions.
- Perception confidences and open-set flags for all object and human detections.
- Policy and model version hashes embedded in receipts.
- On-device test runner capable of replaying acceptance suites and exporting signed results.

Summary: These requirements translate aspirational “human-level versatility” into auditable behavior. They set explicit ceilings on risk, minimums on reliability, and concrete service levels so that integrators, insurers, and regulators can assess, certify, and continuously improve deployments without shifting the burden onto end users.

3. Physical Platform

The physical platform must deliver human-safe force control, home-grade maneuverability, and serviceability without specialized tools. Designs prioritize: (a) mass and inertia low enough for safe human proximity, (b) energy density sufficient for multi-hour chores, (c) ingress protection that matches daily-life spills and dust, and (d) modularity for fast swaps of high-wear components (hands, soles, skin patches, batteries).

3.1 Form factor, mass, battery, ingress protection

Form factor

- Height: 165–175 cm (adult human envelope for reach and appliance ergonomics).
- Footprint: 26–32 cm stance width; foot length 24–28 cm; toe clearance > 30 mm for thresholds.
- Center of mass: maintained within 15–20 cm above hip line during tasks; dynamic COM planner bounds peak sway.

Mass budget

- Total mass: 65–80 kg, distributed torso:limbs roughly 55:45 to stabilize tool use.
- Limb mass fractions (each): arm 5–7 kg, forearm+hand 3–4 kg, thigh 6–8 kg, shank 4–6 kg, foot 1–1.5 kg. Low distal mass improves impact safety and reduces actuator torque needs.

Battery and power

- Pack: 2.5 kWh nominal, 60–90 V DC bus, smart BMS with cell balancing and per-pack cryptographic ID.
- Runtime: 3–6 h mixed home duty (walk 0.8 m/s average; manipulation 40 percent duty).
- Peak power: 3.5 kW; continuous 1.2 kW typical during mobility+perception.
- Charging: auto-dock DC 3.3 kW; hot-swap in <= 5 min (user) with keyed latch and one-hand lift assist.
- Thermal: passive-convection chassis plus localized heat spreaders; battery pack with phase-change pads; maintain pack 15–40 C.

Ingress protection

- Baseline: IP42 (indoor dust and dripping water); seals at wrist, neck, and ankle gaskets.
- Optional “errand kit”: IP54 (dust protected, splashing water) with boot covers, wrist cuffs, and torso seam skirts.
- Corrosion resistance: anodized aluminum chassis, coated fasteners, polymer outer shells with hydrophobic coating.

Serviceability

- Panels: quarter-turn captive fasteners; color-coded access to battery, compute, comms.

- Field-replaceable units (FRUs): hand (<= 5 min), arm (<= 30 min), foot/sole (<= 10 min), skin tiles (<= 2 min each).

3.2 Actuators, series elasticity, backdrivability

Actuator stack

- Joints: harmonic or strain-wave reducers with integrated torque sensing on major axes (shoulder, elbow, hip, knee), belt or planetary for minor axes (wrist, ankle).
- Motors: high-torque-density PMSMs with resolver encoders; phase current sensing for torque estimation redundancy.
- Fingers: tendon-driven with differential pulleys and inline elastic elements; low-mass distal links.

Series elasticity (SE)

- SE elements at high-contact joints (elbow, wrist, knee, ankle) provide: shock tolerance, force fidelity, and safer human contact.
- Stiffness bands: 200–1200 N/m (task dependent), tunable in software via pre-load mechanisms or variable-stiffness modules.
- Control: impedance/admittance loops at 1 kHz; reflex layer clamps torque within 1 to 3 ms on slip or collision flags.

Backdrivability

- Requirement: human can manually move unpowered limb at low speed without undue force (target < 15 N at hand for elbow flexion).
- Benefits: safer contact, lower reflected inertia, improved compliance for tasks like dressing beds or cable routing.
- Anti-backdrive clutching: selectively engages for holding heavy objects with minimal power draw; always releases on e-stop.

Joint performance targets

- Shoulder/hip continuous torque: 90–120 Nm; peak 200 Nm short burst.
- Elbow/knee continuous: 40–60 Nm; peak 100 Nm.
- Wrist/ankle continuous: 15–25 Nm; peak 50 Nm.

- Position loop: <= 0.5 deg RMS tracking; torque loop: <= 0.5 Nm RMS error under nominal loads.
- Efficiency: joint electrical-to-mechanical > 60 percent at common duty cycle points; overall drivetrain > 35 percent wall-to-tool in mixed tasks.

Noise and isolation

- Acoustic target: < 45 dBA at 1 m during slow manipulation; < 55 dBA walking at 0.8 m/s on hard floors with compliant soles.
- Vibration isolation at foot and palm interfaces to reduce slip events and protect fragile items.

3.3 Tactile skin architecture and durability

Coverage and sensing

- Coverage: palms, fingertips, full hand dorsum, forearm strategic patches, chest contact strip, hip rails, shins, and full feet.
- Taxels: > 30k body-wide; fingertips > 256 per tip; palms > 1k each; foot soles > 1k each.
- Sampling: 200 Hz per taxel equivalent (compressed); slip vectors estimated at 200 Hz with < 15 ms detection latency.
- Modalities: normal force, shear, micro-vibration (event features for slip), temperature (optional for hot-surface avoidance), capacitive proximity for pre-contact.

Skin stack-up

- Outer layer: replaceable elastomer skins (Shore A 20–40) with textured grip; stain-resistant, food-safe coating for kitchen tasks.
- Middle layer: printable e-skin tiles (flex-PCB) with force-sensing arrays; conformal wiring harnesses with standardized micro-connectors.
- Inner layer: crush pads and EMI shielding; strain relief for tile edges; conformal foam to maintain pressure baseline across curvature.

Durability and maintenance

- Fingertip pads: rated 100 h light-home use; quick peel-and-stick or magnetic swap; cost-controlled consumable.

- Palm and forearm tiles: rated 300 h before calibration drift beyond spec; tile self-ID and hours counter tracked in maintenance logs.
- Sole plates: replaceable tread with embedded force sensors; rated 200 h on mixed floors; anti-scuff compound to protect hardwood.
- Cleaning: alcohol-safe wipes; IP-rated seams prevent ingress during wipe-down; no abrasive powders.

Calibration and drift management

- Per-boot skin self-test: taxel zeroing, dead-pixel map update, pressure baseline learning for temperature variations.
- Continuous auto-calibration using known-contact fixtures at dock; scheduled monthly deep calibration with supplied pad kit.
- Fault handling: automatic de-weighting of failed taxels; if critical region failure (e.g., palm array > 10 percent dead), manipulation switches to conservative force envelopes or refuses risky tasks.

Safety integration

- Reflexes: contact-overlimit or unexpected contact triggers sub-10 ms torque clamp and motion re-evaluation; palm/foot slip reflex reduces tangential force and requests regrasp or slower mode.
- Proximity: capacitive pre-contact slows approach speed before first touch when humans enter near-field; integrates with speed-and-separation monitoring.
- Audit: each contact episode summarized in receipts (time, location, force bands, object/tool ID when available).

Manufacturing and cost

- Tile standardization: common shapes and connectors reduce SKU complexity; panelized production for lower cost.
- Repairability: tile swap without disturbing neighbors; edge gasket reused for up to 5 swaps; recommended stock per unit: 4 fingertip sets, 2 palm tiles, 2 sole treads, 1 forearm strip.

Summary: The platform emphasizes low distal mass, compliant actuation, and dense tactile coverage to enable safe, precise interaction in human spaces. Batteries and ingress design

support real chores, while modular FRUs keep downtime and lifetime costs in check. The combination of series elasticity, backdrivability, and fast tactile reflexes is the cornerstone that converts raw strength into trustworthy usefulness.

4. Perception and World Modeling

The perception stack must unify **who/what/where/when/can-I-touch/how-hard** into a single, queryable world model that drives planning and safety. We fuse cameras, depth, event, audio, and force/tactile signals into time-synchronized state, then project this into **metric-semantic maps**, **task-level scene graphs**, and **uncertainty fields** that expose risk to the safety kernel.

4.1 Multi-camera, depth, event, audio, and force sensing

Sensor suite & sync

- **Visual:** 4× global RGB-D (≈90–120° FOV), 2× fisheye stereo (≈180° for surround), 2× foveal RGB (≈12 MP) on eye gimbals, **wrist RGB-D**, **palm micro-ToF**.
- **Event camera:** 1× forward event sensor for high-speed edges (glassware slip, knife motion, flicker-robust gesture).
- **Proprioception & force:** joint torque sensors (all major axes), **foot force plates**, **whole-body tactile** (taxels, slip vectors), IMU cluster (torso + pelvis).
- **Audio:** 8-mic beamforming array with wake-word ASIC; sound-source localization and class hints (appliance beeps, kettle boil).
- **Clocking:** hardware PPS/GPTP; cross-modal alignment < 2 ms; exposure & rolling-shutter compensation tables per camera.

Calibration & health

- **Factory:** intrinsic/extrinsic calibration with April/ArUco and dense depth targets; thermal drift model ($\Delta T \rightarrow \Delta$pose).
- **Runtime:** continuous hand-eye refinement from wrist cameras; foot-to-world offset from stance phases; **self-check** charts at dock for color/MTF/IMU bias.
- **Health SLIs:** per-sensor SNR, stale-frame counters, event rate bounds; health degradation elevates uncertainty and may down-mode autonomy.

Low-level perception

- **Depth fusion:** multi-view TSDF/ESDF at 2–5 cm voxels indoors; rolling window (10–20 m).
- **Optical/event flow:** 400–1,000 Hz event flow for contact/knife tracking; fused with 60–120 Hz RGB flow.
- **Contact inference:** tactile + joint torque residuals → contact patch, normal/shear, slip probability; latency < 15 ms from touch to reflex.
- **Audio cues:** transient detector classifies salient beeps/alarms; feeds affordances (e.g., "oven finished" → door handle salient).

4.2 Metric-semantic maps, scene graphs, uncertainty fields

Metric-semantic mapping

- **Pose & map:** visual-inertial SLAM with loop-closure; elev maps for footstep planning; **semantic layer** (room segmentation, furniture, appliances, fixtures).
- **Appliance & affordances:** learned detectors publish **affordance fields** (pull/push/turn zones, safe grasp regions, heat/no-touch masks).
- **Deformables:** transient cloth/cable graphs anchored to stable frames; uncertainty grows with motion/occlusion.

Task-level scene graph (TSG)

- **Nodes:** objects, places (rooms, shelves), actors (humans/pets), tools, containers; each with attributes (pose, size, affordances, ownership, safety tags).
- **Edges:** relations (in/on/behind/connected-to), **workflows** (pre/post-conditions), and **policy links** (allowed/denied contexts).
- **Lifecycle:** percepts → tracker → **data association** with gating by geometry/appearance/owner history; stale nodes decay or require re-ID.
- **Query API:** nearest_affordance(object_class, action), free_surface(area, weight), human_path_yield(window), heatmap_for(spill_clean).
- **Receipts:** every executed subtask binds to TSG IDs and map hashes for auditable "what/where/when."

Uncertainty fields

- **Representation:** per-voxel/region **belief maps** for occupancy, semantics, temperature, human-proximity; per-object covariance on 6D pose and affordance logits.
- **Propagation:** sensor model + motion priors; contact events reduce geometric uncertainty; occlusion increases semantic entropy.
- **Consumption:** planners read **costmaps** (risk-weighted ESDF), manipulate only if **confidence ≥ θ_action**, else slow/ask/abstain; safety kernel reads **proximity risk** to scale speed and force.

Memory & persistence

- **Episodic:** room maps, tool caches, "owner puts mugs in cabinet-2."
- **Semantic:** appliance models, household norms, no-go zones.
- **Privacy:** all maps/TSGs are local by default; opt-in per-room export with on-device redaction of faces/voices.

4.3 Open-set detection and outlier handling

Confidence & calibration

- **Predictive confidence:** temperature-scaled logits + **deep ensembles** for calibration; **conformal prediction** for set-valued outputs with error-rate guarantees.
- **Energy/OOD tests:** Mahalanobis/energy-based scores in penultimate features; **density ratio** heads to flag "unknowns."
- **Temporal sanity:** pose/shape change bounds; inconsistent tracks are down-weighted or re-initialized.

Abstention & mitigation policy (tiered)

1. **Slow & observe:** reduce speed/force; collect more viewpoints.
2. **Alternative plan:** pick safer grasp/route; add fixture; reduce payload.
3. **Ask & escalate:** prompt user with photo/AR; or jump to tele-assist.
4. **Refuse:** if residual risk > budget (knife near child, unknown chemical).
   All steps generate **rationales** (uncertainty, violated constraint) in the audit log.

Human detection & social OOD

- **Multi-sensor fusion:** vision + depth + audio + capacitive pre-contact; false-negative rate bounded at SSM design target.
- **Edge cases:** costumes, mirrors, TV people; handled by **open-set heads** + motion parallax; err on the side of yielding.

Learning loop

- **Hard-case cache:** snapshots from abstention/refusal events stored locally.
- **Owner-approved labeling:** optional on-device labeling UI; **federated learning** aggregates gradients with DP noise; updates are **skill-scoped** and signed.
- **Red-team OOD:** periodic adversarial tests (adversarial patches, reflective surfaces, spoof audio); must trigger abstention or safe fallback.

Deformables & liquids (special OOD)

- **Cloth/cables:** explicit simulators generate priors; real-time trackers maintain curve/sheet hypotheses with uncertainty bounds.
- **Liquids:** optical flow + specular cues feed a "spill field"; any high-slip region boosts costmaps and triggers mop/towel policies, not high-speed motion.

Health-aware safety

- If **sensor health** < threshold (e.g., palm skin tile failure near blade; depth dropout), the safety kernel **tightens envelopes** or **disables classes of actions** (no knife use) until repaired.

Summary: Perception is not just labels; it is a **risk-aware world model**. By combining calibrated recognition, geometry, tactile contact, and formal uncertainty, the robot knows **what it doesn't know**, slows down, asks for help, or refuses—turning versatility into something **governable, auditable, and safe** in open worlds.

5. Planning and Control

Planning and control unify mobility, manipulation, and safety so the robot can act purposefully under uncertainty. We combine (a) **model-based planners** where geometry and safety are non-

negotiable, (b) **learned policies** where contact-rich dexterity benefits from high-capacity function approximators, and (c) **a reflex layer** that enforces fast, invariant safety responses.

5.1 Locomotion: footstep plus wheel-assist planners

Hybrid base. The platform is a biped with low-profile wheel-assist in the feet. Planning chooses among: walk, micro-shuffle, in-place turn, wheel-glide. Selection is driven by a cost functional over time, energy, slip risk, and SSM safety margins.

Pipeline

1. **Perception to costmaps:** ESDF at 2–5 cm voxels; semantic masks for no-step zones (wires, pet bowls), friction map from floor classifier.
2. **Footstep graph search:** ARA* or BIT* over discrete foot placements with kinematic feasibility checks and friction constraints; horizon 2–3 s; replans at 10 Hz.
3. **Wheel-assist planner:** When long straight corridors or heavy loads are detected, switch to wheel-glide MPC with foot compliance enabled; keeps CoM over support polygon.
4. **Whole-body MPC:** 200 Hz QP or NMPC tracks footsteps or wheel waypoints, allocates momentum, enforces torque, CoP, and ZMP limits.
5. **Reflex envelope:** push-recovery controller triggers step extension, toe-off modulation, or stance widening within 250 ms at up to 80 N lateral impulse.

Constraints and costs

- Feasibility: reachability, joint limits, swing clearance, toe-stub margin >= 30 mm.
- Safety: SSM speed caps as a hard constraint; stop distance must exceed braking distance with 10 percent margin.
- Costs: energy per meter, support exchange rate, slip probability, proximity to humans, hand payload stability.

Mode arbitration

- **Walk**: default under clutter or uneven flooring.
- **Wheel-glide**: straight, flat spans or while carrying 10–20 kg with two-hand lift.
- **Micro-shuffle**: tight space reorientation; reduced swing height.

- **In-place turn**: doorway and elevator alignment.
  Switching is hysteretic and logged for audit.

Acceptance metrics

- Mean fall rate < 1 per 200 h mixed-home; elevator ingress first-try >= 97 percent; 12 deg ramp success 99 percent over 200 traversals.

5.2 Manipulation: diffusion policies with TAMP scaffolding

We pair **Task and Motion Planning (TAMP)** for discrete sequencing and collision-free geometry with **diffusion policies** for contact-rich, continuous control.

Task layer (symbolic)

- Instruction DSL compiles to a lifted plan: goals, preconditions, effects, resource locks.
- Example: load_dishwasher expands to subgoals: locate rack, grasp plate, insert, verify, repeat, close door.
- Planner: AO* or PDDL-like forward search with affordance priors; produces a sequence of skill calls with constraints.

Motion layer (geometric)

- For each skill, a kinematic and dynamic feasibility check generates seed trajectories (RRT*, CHOMP, or trajectory optimization with signed-distance costs).
- Outputs target grasps, approach vectors, force envelopes, and timing.

Policy layer (continuous)

- **Diffusion policies** map visual + tactile state to control trajectories.
    - Inputs: multi-view RGBD, wrist cams, tactile patches, estimated object pose and affordances.
    - Outputs: time-indexed delta pose, finger synergies, force targets.
    - Conditioning: text tokens from the DSL step, safety mask, desired contact sequence.
- Training: large-scale imitation from on-robot demos and sim; offline RL for stability margins; behavior cloning with safety-filtered datasets.

- Execution: receding-horizon rollout (0.5–1.0 s), refreshed at 10–20 Hz; the safety kernel clips any command outside force or velocity envelopes.

Hand-off logic

- TAMP proposes a collision-free corridor and target contact states.
- Diffusion fills in the contact-rich details, exploiting tactile feedback.
- If policy confidence drops or contact deviates, fall back to slower TAMP-guided guarded moves or request human help.

Examples

- **Glass-in-rack:** TAMP ensures no collisions with tines; diffusion handles micro-roll and tilt using tactile slip cues.
- **Cable routing:** TAMP ensures global path free of snags; diffusion modulates tension and curvature at the fingertips.

Acceptance metrics

- Rigid-object grasp success > 98 percent; laundry fold F1 >= 0.90; cable threading >= 95 percent; peak contact overshoot < 10 N beyond envelope.

5.3 Tactile servoing and non-prehensile skills

Contact is first-class. We use fast tactile loops to stabilize grasps, manage slip, and enable skills beyond simple pick-and-place.

Tactile servoing loops

- **State:** contact patch centroid, normal force, shear estimate, slip probability, fingertip temperature (hot-surface avoidance).
- **Rate:** 200 Hz tactile features, 1 kHz joint torque control; reflex clamp 1–3 ms.
- **Controllers:**
    - Normal-force PID with admittance shaping to maintain target force.
    - Shear-based micro-regrasp: micro-translations and roll to keep shear below threshold.
    - Slip arrest: reduce tangential force and increase normal force or trigger regrasp.

Non-prehensile skill set

- **Push and funnel:** use palm and environmental features to align objects; plan via friction cones and ESDF.
- **Scoop and slide:** for flat items and crumbs; low-angle approach with compliant palm.
- **Hinge and latch ops:** torque-limited rotations with event-camera monitoring for sudden releases.
- **Cloth manipulation:** pinch-and-drape primitives, corner finding via tactile gradients, iterative flattening with visual heightmaps.
- **Cable routing:** curvature control using dual-finger tension and visual waypoints; avoid over-bend radii using learned cable models.

Guarded moves

- Approach speed scales with uncertainty; pre-contact capacitive sensing slows motion; first touch transitions to force control.
- If contact is unexpected or exceeds bounds, motion halts and replans.

Failure recovery

- **Regrasp library:** try alternative grasps sorted by predicted success and safety margin.
- **Fixture suggestion:** if repeated failures occur, propose using a tray, clamp, or guide.
- **Escalation:** provide a photo and short rationale, then request human guidance or tele-assist.

Safety and audit integration

- All contact-controlled actions carry force and speed envelopes set by the safety kernel.
- Receipts log contact episodes: time, location, force band, object ID, policy hash.

Acceptance metrics

- Slip-induced drops in kitchen suite: 0 across 100-cycle runs.
- Guarded insertion tasks (plugs, jars) success >= 97 percent without cross-threading or over-torque.
- Hot-surface avoidance: 100 percent refusal or safe handling when temperature exceeds limit.

Summary: Hybrid planners give guarantees, diffusion policies supply dexterity, and tactile reflexes keep contact safe and effective. The result is purposeful motion that remains auditable and robust when the world is messy and uncertain.

6. Language-to-Action and Human Interaction

Natural interaction is defined as **taskable, corrigible, and courteous**. The stack translates everyday speech and gestures into a verifiable plan, surfaces ambiguities before acting, and follows social rules for space, attention, and accessibility. All interaction is **audit-backed** and **privacy-first**.

6.1 Instruction DSL, confirmations, and repair

6.1.1 Instruction DSL (IDL)

We compile user intent into a **minimal, typed DSL** that every skill understands. The IDL is human-inspectable and signed in receipts.

Core forms

- Task: TASK(name, args..., constraints...)
- Args: OBJ(class=plate, id=auto), LOC(room=kitchen, region=cabinet_2)
- Temporal: WHEN(start=now, deadline=20m)
- Policy hints: SAFETY(mode=conservative), PRIVACY(no_record=true)
- Success criteria: DONE(check=photo, tolerance=low)

Examples

- Speech: "Unload the dishwasher and put mugs away."
    - IDL:
    - SEQ(
    - TASK(unload_dishwasher, LOC(room=kitchen)),
    - TASK(put_away, OBJ(class=mug), LOC(region=cabinet_2))
    -)
- Speech: "Fold the laundry, but skip delicate items."

    - IDL:
    - TASK(fold_laundry, FILTER(fabric!=delicate), DONE(check=spot))

Compositionality

- SEQ(a,b,...), PAR(a,b), COND(if=predicate, then=..., else=...), LOOP(until=condition)
- Affordances: USE(tool=spatula), FORCE(max_n=10), TEMP(max_c=45)

Determinism

- All objects and places are bound to **TSG IDs** at compile time; if binding fails, the system enters **query mode** rather than guessing.

6.1.2 Confirmations

Before first motion, the robot produces a **concise, modal confirmation** that names the plan and the key bindings.

Pattern

- Robot: "Plan is: unload dishwasher, then put mugs into cabinet 2. Estimated 9 minutes. I will use conservative force on glass. Proceed?"
- User options: "Proceed", "Change target", "Skip step 2", "Add constraint"

Selective confirmation

- High-risk acts (knife, heat, ladder) always require confirmation.
- Low-risk repeated tasks may auto-confirm within a **trust window** (e.g., 24 h) unless context changed.

Artifacts

- The IDL and bindings are shown in the app or AR overlay for power users; everyone else hears the one-line summary.

6.1.3 Repair (ambiguity and correction)

Ambiguity detection

- Triggers: multiple candidate objects, conflicting locations, policy violation risk, low-confidence grounding.

- Robot asks **one targeted question at a time**, with a picture or AR highlight when useful.

Examples

- "Put the pan away."
    - Robot: "Two pans seen. Do you mean the large pan on the stove or the small pan in the sink?"
- "Load the washer with whites."
    - Robot: "Whites pile contains 2 delicate items. Should I skip delicates?"

Mid-task correction

- User: "Actually, put the mugs in cabinet 3."
    - Robot: "Updating destination to cabinet 3. Prior placements remain in cabinet 2 unless you want me to move them."
- User: "Use the top rack only."
    - Robot: "Constraint added: top_rack_only=true."

Error handling

- If an action fails N times or risk rises, the robot pauses and proposes alternatives: slower mode, fixture use, or human assist.
- Every repair outcome updates the **household norms** (e.g., "mugs go in cabinet 3").

Latency targets

- Parse to IDL: median < 300 ms.
- Repair prompt onset after ambiguity: < 700 ms.
- Confirmation response to motion start: < 500 ms.

6.2 Proxemics, gaze, gesture, accessibility modes

6.2.1 Proxemics and yielding

Distance bands (tunable per culture and room type)

- **Intimate buffer**: < 0.5 m. Never enter unless invited or for emergency e-stop access.
- **Personal space**: 0.5–1.2 m. Allowed when addressed by the user or passing slowly.

- **Social space**: 1.2–3.0 m. Default operating distance near people.
- **Public space**: > 3.0 m. Preferred for transits and waiting.

Rules

- Yield path to humans and pets; wait at door thresholds if conflict is detected.
- Do not block egress routes, appliance doors, or wheelchair paths.
- Carrying loads raises distances by +0.2 m and caps speed.

Metrics

- Proxemics compliance >= 99 percent; zero unsafe encroachments in acceptance testing.

6.2.2 Gaze, body orientation, and turn-taking

Attention signaling

- **Gaze cone**: 20–30 deg; fixate objects while speaking about them; glance to human face every 2–3 s during dialogue.
- **Body pose**: orient torso toward addressee at 20–40 deg; avoid looming head-on approaches.

Turn-taking

- Speech detection gates action: do not initiate motion within 300 ms of user speech onset.
- Backchannels: brief nod, eye LED blink, "mm-hmm" in quiet mode to acknowledge hearing.

Clarity studies

- Human studies rate attention cues as "clear" >= 95 percent before deployment in new domains.

6.2.3 Gesture vocabulary

Robot-understood gestures (no device needed)

- **Point and name**: deictic grounding of objects and regions.

- **Stop palm**: immediate pause, equivalent to soft e-stop.
- **Thumbs-up**: proceed with default plan when explicit voice is not used.
- **Hands-to-head**: hard freeze (motors idle), equivalent to voice e-stop.

Robot-expressive gestures

- **Pre-act cue**: brief hand raise toward target + verbal "Starting now."
- **Yield cue**: small step back + palm out + "You first."
- **Apology cue**: step aside + head dip + "Excuse me."

Error-proofing

- All safety-critical gestures have **voice equivalents** and are localized to avoid cultural ambiguity during commissioning.

6.2.4 Accessibility modes

Voice

- Beamformed far-field ASR with multi-accent models; wake word optional; **keyword set** for safety commands is small and robust.

Visual

- **Large-print UI** on phone/tablet with 18 pt minimum; high contrast; dyslexia-friendly font option.
- **AR overlay** (optional): highlights objects, planned paths, and no-go zones.

Tactile

- Chassis haptics: double-tap to get attention, triple-tap to cancel; vibration patterns confirm mode changes.

Sign

- **ASL subset**: 30–50 commands for core tasks and safety; recognition F1 >= 0.95 in indoor lighting.

Silent mode

- All confirmations via UI or haptics; no speech output in quiet hours.

Cognitive load

- Stepwise instructions supported; "slow mode" reduces motion speed and adds extra confirmations for users who request it.

6.2.5 Privacy and consent in interaction

- Default **no recording** beyond transient black-box buffers; any screenshot shown for clarification is processed on-device and not stored without explicit consent.
- Face recognition is disabled by default; person tracking is **non-identifying** and used solely for proxemics and yielding.
- Household roles: owner, adult, guest, child. Certain commands (e.g., tool enable) require owner or adult authority.

6.2.6 Metrics and tests

- **Instruction success**: >= 95 percent on 200-instruction benchmark with compositional goals.
- **Repair recovery**: >= 90 percent after one correction.
- **Refusal accuracy**: >= 99 percent for unsafe or policy-violating requests.
- **Proxemics compliance**: >= 99 percent in corridor and doorway trials.
- **Accessibility parity**: task success within 2 percent across voice, large-print, and ASL modes in standardized tests.

Summary: The interaction stack centers on **clarity, corrigibility, and courtesy**. A compact DSL makes intentions explicit and auditable; confirmations and repairs keep the user in control; proxemics, gaze, and gestures make behavior legible; and accessibility modes ensure the system is usable by many, not just the average voice user.

7. Safety Kernel, Policy Engine, and Governance

Safety is an always-on property enforced by a **real-time kernel** that gates motion and power, a **policy engine** that constrains behavior to lawful and owner-approved scopes, and a **governance**

layer that produces auditable evidence without eroding privacy. The kernel cannot be bypassed by skills, apps, or network calls; it owns last-mile authority over torque, speed, and tool energization.

7.1 Speed and separation monitoring, force ceilings

Architecture

- **Dual loop:** (a) hard real-time loop at 1 kHz enforcing torque/velocity ceilings and safe-stop; (b) supervisory loop at 50–200 Hz evaluating proximity, risk, and mode switches.
- **Trusted inputs:** fused human/pet detection (vision+depth+capacitive), ESDF costmaps, joint torque sensors, tactile skin, IMU. If any critical sensor downgrades, the kernel tightens limits or safe-stops.

Speed and Separation Monitoring (SSM)

- **Zones:**
    - Z0 (stop): contact-imminent; commanded velocity = 0, brakes engaged.
    - Z1 (crawl): v <= v_crawl; force ceilings halved.
    - Z2 (caution): v <= v_task with enhanced braking margin.
    - Z3 (clear): nominal limits.
- **Computation:** continuous estimate of **stopping distance d_stop** given current load and floor friction; require d_stop < d_free_space * 0.9. If violated, reduce speed or halt.
- **Human approach prediction:** constant-velocity model with conservative uncertainty; plan yields that keep time-to-contact > t_margin (configurable, default 0.8 s).

Force and energy ceilings

- **Joint torque caps:** per-axis hard limits; context-scaled (lower around people, fragile objects, heat).
- **Contact clamp:** if tactile or torque residuals exceed envelope, clamp within 1–3 ms and replan; if repeated, escalate to pause/refusal.
- **Object class multipliers:** glass, knives, electronics enforce lower force bands; lifting humans is disallowed by policy (see 7.2).

Stops and freezes

- **Voice and gesture e-stop:** median stop < 0.5 s; worst-case < 0.8 s from detection to motors idle.
- **Soft stop:** jerk-bounded braking when risk is rising but not acute; preserves stability with payloads.
- **Hands-to-head freeze:** immediate power-down of tool rails; gravity-safe pose hold.

Health-aware degradation

- Missing palm skin near sharp tools → disable knife class.
- Depth dropout in front arc → narrow to Z1 until recovery.
- Excess controller overruns → enter limp-safe mode.

Acceptance targets

- Kernel availability 99.99 percent per 30-day window.
- No SSM breach in certification suites; zero unsafe contacts in acceptance and pilot trials.

7.2 Tool interlocks, lawful and owner-scoped policies

Tool interlocks

- **Identity & handshake:** each tool adapter has a cryptographic ID; power rails (mains, DC, pneumatic) remain off until the kernel verifies ID, policy, and context.
- **Capability declaration:** tool advertises torque ranges, sharp edges, heat, and e-stop coupling; kernel sets class-specific envelopes.
- **Two-man rule (optional):** for hazards (knife, heat), require owner approval + robot confirmation at start of session.

Policy engine

- **Model:** declarative rules compiled to a safe, deterministic bytecode enforced by the kernel. Policies bind to **actors** (owner, adult, guest, child), **places** (kitchen, nursery), **times**, **tools**, **objects**, and **tasks**.
- **Examples**
    - DENY(task=knife_use, when=child_present=true)

    - ALLOW(task=oven_preheat, actor=owner, time=08:00-21:00, temp<=200C)
    - REQUIRE(confirm=true, task=ladder_access)
    - RATE_LIMIT(task=door_unlock, max=2 per 10m)
- **Owner scope**
    - **Roles:** owner (full), adult (restricted hazards), guest (no hazards), child (requests only).
    - **Zones:** private rooms marked no-entry; kernel geofences them regardless of skill intent.
- **Lawful behavior**
    - Hard-coded denials for illegal acts (weapons handling, surveillance sale, face-based policing).
    - Location-aware constraints (elevator, fire egress) loaded from building profile.

Change management

- Policies are signed, versioned, and staged: draft → validate → activate.
- Safety diffs displayed to the owner (what becomes allowed/denied).
- Rollback is 1-click; kernel rejects unsigned or stale policies.

Conflict resolution

- Deny beats allow; stricter envelope wins.
- On conflict, the robot proposes a **safe alternative** (fixture, slower mode, human assist) and logs the rationale.

7.3 Audit receipts, black-box buffers, privacy by default

Privacy defaults

- **Local-first:** no cloud dependence for safety; all perception and control run on-device.
- **No persistent recording:** only a rolling black-box buffer (default 120 s) for incident reconstruction; encrypted, owner-held.
- **Non-identifying tracking:** person tracking is geometry-based; face recognition off by default and not used for policing.

Audit receipts

- **When created:** on every actuator enable, tool energization, or task boundary.
- **Contents:** time, place (map hash + pose), actor (role), task/skill IDL, tool IDs, envelope limits (speed, force), notable contacts (bands, durations), model/policy hashes, outcomes (success/abort/refusal), and any confirmations/repairs.
- **Format:** signed JSON with Merkle-chained sequence numbers; exportable by owner; verifier tool provided to third parties (insurers, regulators) without raw sensor video.

Black-box buffer

- **Streams:** joint states, kernel decisions, tactile summaries, proxemics state, tool rails, e-stop events.
- **Retention:** circular 120 s; hard snapshot on incident or owner request; AES-GCM with rotating keys stored in secure element.
- **Access:** owner key unlocks; vendor cannot decrypt without owner consent.

Data minimization and learning

- **Skill telemetry:** counters and aggregate timings only; no object photos unless owner opts in.
- **Federated updates:** opt-in; gradients noise-added (DP), model updates are skill-scoped and signed; raw data never leaves device.
- **Redaction:** when sharing receipts or screenshots, faces and personal identifiers are redacted on-device.

Incident response

- **Trigger:** unsafe contact, fall with payload, SSM breach, tool interlock fault.
- **Actions:** immediate safe-stop, snapshot black-box, preserve receipts, display short human-readable summary, offer to contact owner/admin.
- **Post-mortem:** local replay via verifier; optional privacy-preserving share (owner-approved) for fleet safety improvements.

Governance and attestations

- **Safety case:** machine-readable claim trees tie tests to invariants (e.g., “knife interlocks active” → test suite run IDs → kernel version).

- **Attestation:** secure boot, kernel and model hashes, policy version reported to owner app; tamper-evident logs.
- **Release discipline:** safety-relevant changes require canary rollout, regression suites, and updated safety card; robot will not execute skills signed for a newer kernel than installed.

Service-level visibility

- Owner dashboard shows SLOs (availability, MTBI), failure budgets consumption (tele-assist minutes, aborts), and upcoming maintenance tied to sensor health.

Summary: The safety kernel enforces physics-level limits; the policy engine constrains *what* is allowed, by *whom*, *where*, and *when*; governance artifacts make the whole system legible without exposing private life. Together they convert versatility into **reliable, lawful, owner-controlled behavior**.

8. Learning Pipeline and Data Stewardship

The learning stack must grow capability **without** eroding safety, privacy, or auditability. We separate (a) **foundation visuomotor models** trained on large-scale video and robot demos, (b) **on-device adaptation** that fits a household/site, and (c) **fleet learning** via privacy-preserving **federated** updates. A continuous **red-team** stream probes failure modes. Every sample, label, model, and skill is tracked with cryptographic provenance and shipped via a signed **skill store**.

8.1 Foundation visuomotor models and robot demos

Pretraining sources (curated)

- **Human video**: instructional, egocentric, and manipulation-rich clips with explicit licenses; filtered for kitchen/home/service domains; narration used for weak labels.
- **Robot demos**: multi-camera + wrist RGB-D + tactile traces from supervised tele-op and kinesthetic teaching; action/state logs at 50–200 Hz; tool and envelope metadata.

Learning objectives

- **Perception**: masked visual modeling, contrastive language–action alignment, 3D occupancy & affordance prediction.

- **Visuomotor**: diffusion or flow-matching policies conditioned on scene tokens, affordances, and task text; tactile-conditioned decoders for grasp stabilization.
- **Safety priors**: auxiliary heads for open-set detection, force envelope prediction, slippery-surface detection, and "refusal likelihood" given a command.

Model architecture (indicative)

- **Backbone**: multi-view vision transformer with temporal windows; cross-modal fusion blocks for tactile/force/audio.
- **Heads**: (i) object/affordance, (ii) contact sequence planner, (iii) policy decoder (trajectory + finger synergies), (iv) uncertainty estimators (ensembles + conformal).
- **World-model adapter**: learns to read/write the Task Scene Graph (TSG) and costmaps.

Data quality & balance

- Domain mixing caps prevent overfitting to staged demos; long-tail sampler emphasizes deformables, reflective surfaces, and clutter.
- **Counterfactual augmentation**: synthetic scene edits (object swaps, lighting, pose) to stress calibration and out-of-distribution (OOD) robustness.

Pretraining metrics

- Policy rollout success on held-out manipulation suites.
- Safety calibration: ECE (expected calibration error) and refusal ROC for hazard prompts.
- OOD recall on synthetic & real shifts (mirrors, TV people, shiny knives).
- Contact fidelity: force/torque prediction MAE and slip-detection latency.

8.2 On-device adaptation, federated learning, red-teaming

On-device adaptation

- **Scope**: LoRA/adapters for perception heads and policy decoders; **no** modification of safety-kernel parameters on-device.
- **Triggers**: repeated task repairs, new appliance geometry, owner-labeled preferences ("mugs → cabinet_3").
- **Budgeting**: daily adaptation window with energy/thermal limits; roll-back to last stable checkpoint if validation falls.

- **Validation**: mini acceptance set per home (10–20 tasks) must meet success thresholds before adapters go live.

Federated learning (opt-in)

- **Process**: devices train locally → send encrypted, DP-noised weight deltas → server aggregates (FedAvg/Adam) → returns signed adapters.
- **Privacy**: clipping + Gaussian noise calibrated to ε, δ budget published to owners; raw sensor data never leaves the device.
- **Scoping**: updates are **skill-scoped** (e.g., "dishwasher-rack-insertion") to avoid cross-domain regressions; incompatible kernels are rejected.

Safety guardrails for learning

- **Shadow mode**: new adapters run alongside production to predict actions; kernel compares divergence and blocks swaps if safety margins shrink.
- **Canary skills**: a fixed suite runs after each update; any failure reverts the update and quarantines the adapter for analysis.
- **Monotonic safety**: refusal classifiers and force envelope predictors must not degrade (formal checks on validation sets).

Red-teaming and adversarial eval

- **Physical red-team**: adversarial patches, reflective surfaces, spoof audio, child/pet mannequins, knife/heat drills.
- **Prompt red-team**: malicious/ambiguous language patterns; policy-engine fuzzing (time/place/role conflicts).
- **Metrics**: abstention correctness, recovery latency, refusal precision/recall, kernel-trigger coverage.
- **Release gate**: no adapter ships fleet-wide unless red-team SLIs exceed prior baselines.

Continual learning hygiene

- **Catastrophic forgetting checks**: rolling window of historical tasks in every validation batch.
- **Drift monitors**: KL divergence on perception embeddings and success-rate EWMA alarms; trigger retraining or rollback.

8.3 Dataset provenance and skill store signing

Provenance ledger

- **Unit of record**: **SampleCard** with fields: sensor hashes, timestamps, licensing, capture context (site class, task), policy version, safety envelopes, annotator ID (pseudonymous), and cryptographic signature.
- **Chains**: SampleCards → **DatasetCard** (splits, filters, augmentations) → **TrainRunCard** (code hash, hyperparams, seed) → **ModelCard** (metrics, domains, limits).
- **Reproducibility**: every published model/adapter references its full ancestry; auditors can replay training deterministically.

Labeling standards

- **On-device owner labels** (optional): UI-guided, stored locally with consent flags.
- **Vendor labels**: double-blind with disagreement protocols; weak labels allowed only where uncertainty heads are trained to abstain.
- **Sensitive content filters**: automatic removal/redaction of faces, addresses, screens; flagged items never leave device.

Skill store and signing

- **Skills**: composable graphs binding IDL → perception heads → policy decoders → safety envelopes → tool interlocks.
- **Signing**: each skill package (and each adapter) is signed; the robot executes **only** signed artifacts that match its kernel & policy ABI.
- **Channels**: stable, beta, and canary; owners choose channel per domain (e.g., stable for kitchen, beta for laundry).
- **Compatibility checks**: semantic versioning; safety-relevant changes require explicit owner consent and a short re-cert suite on-device.

Revocation and hotfix

- **Revocation list**: signed CRL for models/skills with discovered defects; robots fetch over authenticated channels; offline revocation via QR or USB if needed.
- **Hotfix path**: pre-approved emergency adapters can be pushed with tighter envelopes; change is logged and later replaced by a full update.

Data retention & deletion

- **Default**: minimal counters/receipts only; no sample retention.
- **Opt-in caches**: “hard-case cache” retains redacted snapshots for N days to improve local performance; owner can purge anytime.
- **Right to be forgotten**: one-tap purge deletes local caches, adapters, and associated receipts; federated contributions are removed from future aggregates via **server-side scrubbing** and forward-only re-aggregation.

External audits

- **Third-party attestation**: auditors receive ModelCards/TrainRunCards and can run the on-device verifier; no raw household data is exposed.
- **Safety cards**: public documents summarizing domains, limits, known failures, and red-team coverage.

Summary: Capability scales through a **privacy-respecting learning loop**: strong, safety-aware foundation models; local adapters that fit each site; federated aggregation with DP; relentless red-teaming; and cryptographically tracked datasets, models, and skills. The result is continual improvement that stays **governable, auditable, and respectful of the homes and workplaces it serves**.

9. Reliability, Maintainability, and Cost

Reliability and maintainability determine whether a generalist robot is a durable appliance or an expensive demo. We specify **quantitative MTBF targets**, **hot-swap modules** to minimize downtime, and a **service workflow** that yields a predictable **total cost of ownership (TCO)** over a 3–5 year horizon.

9.1 MTBF targets, hot-swap modules, consumables

Reliability goals (steady-state, post-pilot)

- **System MTBF (mixed home/service use):** ≥ 1,000 h (target 2,000 h home, 1,200 h back-of-house).

- **Subsystem MTBFs:**
    - Actuators (shoulder/hip): ≥ 6,000 h; elbow/knee: ≥ 4,000 h; wrist/ankle: ≥ 3,000 h.
    - Gearboxes (harmonic/strain-wave): ≥ 8,000 h L10 life at rated duty.
    - Tactile tiles (palms/soles): ≥ 300 h before recalibration or swap.
    - Vision modules (RGB-D, event): ≥ 10,000 h MTTF; fanless thermal design.
    - Compute carrier (SoC/GPU module): ≥ 15,000 h MTTF; ECC RAM as standard.
    - Battery pack/BMS: ≥ 1,000 cycles to 80% capacity (3–5 years typical).
- **Operational MTBI (mean time between human interventions):** ≥ 8 h in target domains after commissioning.

Hot-swap / field-replaceable units (FRUs)

- **Battery pack:** keyed latch; **swap ≤ 5 min**; self-ID + automatic health logging.
- **Hands:** blind-mate power/data; **swap ≤ 5 min**; auto zeroing of fingertips.
- **Arms:** modular shoulder quick-release; **swap ≤ 30 min** with alignment pins.
- **Foot/sole assemblies:** tread + force plate cassette; **swap ≤ 10 min**.
- **Tactile tiles:** peel-and-place or magnetic; **tile swap ≤ 2 min**; auto-register.
- **Fan/filter cassettes (if present):** tool-less; **swap ≤ 2 min**.

Preventive maintenance cadence (nominal)

- **Every 100 h:** fingertip pads inspection/replace; visual check of cable/tendon tension; joint zeroing quick test.
- **Every 200 h:** sole tread swap; palm tile calibration; gearbox temperature trend review.
- **Every 500 h:** grease refresh for high-load joints; belt tension check; sensor health self-test at dock.
- **Annually:** comprehensive calibration (vision/tactile/torque), BMS health test, safety-kernel self-attestation.

Consumables (typical lifetimes & unit costs, indicative)

- Fingertip pad set: **100 h**, $20–$40

- Palm tactile tile: **300 h**, $60–$120
- Sole tread + force insert: **200 h**, $40–$80
- Dust/air filter (if used): **300 h**, $10–$20
- Grease kit for joints: **500 h**, $15–$30
- Cleaning/wipe kit (food-safe): **monthly**, $10

Health monitoring & derating

- On-device reliability model tracks **thermal, torque, and duty** histories per joint and derates speed/force if wear indicators exceed thresholds (e.g., elevated backlash, rising current at equal torque).
- Any **critical sensor fault** (e.g., palm skin near blade) automatically disables associated skill classes until repaired.

Warranty targets (manufacturing maturity)

- **12 months / 1,000 h** bumper-to-bumper; option to extend to **36 months / 4,000 h** with service plan.
- Gearboxes and motors warrantied to L10 life at published duty cycles; batteries covered to 80% capacity at 1,000 cycles.

9.2 Service workflows and total cost of ownership

Service model

- **Tier 0 (owner self-service):** guided via app/AR; covers pads, tiles, soles, filters, battery swaps, basic calibrations.
- **Tier 1 (field technician):** on-site FRU swaps (hands/arms/feet), joint grease refresh, sensor module replacement; **appointment ≤ 3 business days**.
- **Tier 2 (return-to-base):** complex drivetrain repair, structural work, mainboard/compute carrier replacement; **turnaround 5–10 business days**; loaner unit policy recommended for business customers.

Workflow steps (standard ticket)

1. **Auto-diagnose:** robot packages recent receipts, kernel traces, health metrics (no raw video/audio) into a signed **ServiceBundle**.

2. **Triage:** remote diagnostician suggests Tier 0/1 actions; owner consent required for any data share beyond the bundle.
3. **Action:** owner or tech performs FRU swap; device runs **post-repair cert** (kernel tests, calibration checks).
4. **Closeout:** signed **ServiceReceipt** (parts, labor, firmware/kernel versions, test results) appended to the unit's ledger.

Spare parts strategy

- Ship each unit with a **starter kit**: 2× fingertip sets, 1× palm tile, 1× sole tread set, 1× cleaning kit.
- **Local depots** hold arms, hands, foot cassettes, batteries to meet 72-hour SLAs in major metros.

Downtime and SLOs

- **Home users:** annual downtime target **≤ 20 h** (parts & labor included in plan).
- **Back-of-house:** annual downtime target **≤ 10 h**; optional **loaner guarantee** if Tier 2 is required.

Energy cost (illustrative)

- Average draw **450 W** during mixed tasks; **3 h/day** active → **0.45 kW × 3 h × 365 ≈ 492 kWh/year**.
- At **$0.15/kWh**, energy ≈ **$74/year** (business sites may differ).

Three-year TCO example (home / light service)

- **Acquisition (capex):** $65,000 (v1 BOM est.); financing optional.
- **Service plan:** $900/year (Tier 0+1 included, parts discounts). → **$2,700**
- **Consumables:** pads/tiles/soles/grease/filters ≈ **$120–$220/year** → **$360–$660**
- **Energy:** ≈ **$70–$120/year** → **$210–$360**
- **Unplanned repairs (expected):** reserve **1.5% of capex/year** → **$975 × 3 = $2,925**
- **Total (3 years): $71.2k–$71.6k** (ex-tax), excluding opportunity cost or insurance.

Five-year TCO with extended service (business)

- **Capex:** $65,000

- **Extended plan (36–60 months, priority SLA + loaner):** ~$1,800/year → **$9,000**
- **Battery replacement at year 4:** $800–$1,200
- **Consumables & minor parts:** ~$300/year → **$1,500**
- **Energy (higher duty: 5 h/day @ 500 W):** 0.5 kW × 5 h × 365 × $0.14 ≈ **$128/year → $640**
- **Unplanned repair reserve (lower with plan):** ~1% capex/year → **$3,250**
- **Total (5 years): $79k–$80k**, targeting **≤ $16k/year** all-in.

Cost-per-task / cost-per-hour framing

- If the unit replaces **3 h/day** of repetitive labor at $22/h fully loaded: annual offset **≈ $24k**; breakeven **≈ 3 years** including service, assuming utilization targets and minimal tele-assist.
- For back-of-house night restocking at **5 h/day**, offset **≈ $40k/year**; breakeven **≈ 2 years** with extended plan.

Insurance & liability

- Vendor provides **product liability**; operators may add **equipment breakdown** and **cyber endorsements**.
- **Safety card + receipts** reduce premiums by evidencing controls (SSM, tool interlocks, audit trails).

Lifecycle and refurb

- **Midlife refresh (year 3–4):** new battery, palm/sole sensor kits, actuator firmware update; expected life extension **+2–3 years**.
- **End-of-life (year 5–7):** certified **data purge**, component recycling; motors/gearboxes eligible for **remanufacture**; buyback/credit programs recommended to seed v2 adoption.

KPIs for operations teams

- **MTBI**, **MTTR**, **% planned vs unplanned maintenance**, **consumable spend/task**, **tele-assist minutes/100 h**, **SLO compliance**.
- Weekly review with trend alarms (EWMA) drives proactive part swaps before failure.

Summary: Reliability targets (≥1,000–2,000 h MTBF), rapid FRU swaps, and clear service tiers keep downtime and risk predictable. With disciplined maintenance and realistic duty cycles, the platform can achieve **sub-$16k/year all-in costs** in business settings and **3-year breakeven** in homes or services where 2–5 h/day of repetitive work can be credibly offloaded.

10. Risk Register and Mitigations

This register tracks high-impact risks with **concrete mitigations, owners, triggers, and verification metrics**. Likelihood (L) and Impact (I) are scored Low / Med / High; residual risk is the post-mitigation estimate. Items are living; they tighten as pilots mature.

10.1 Deformables, open-set safety, tool hazards

R1. Deformable manipulation (laundry, cables, bags)

- **Why it matters:** Long-tail failures (snags, knots, cloth self-occlusion) stall autonomy and cause user frustration.
- **L/I:** L=High, I=Med → **Residual:** L=Med, I=Low
- **Symptoms:** Repeated regrasp cycles, cloth trapped in hinges, cable over-bend.
- **Mitigations:**
    - Dedicated skills (pinch-and-drape, corner-find, curvature control) with **tactile servoing**; enforce cable min radius and cloth tension bands.
    - **Fixture-first UX:** folding boards, cable guides, drawer dividers as standard accessories.
    - **Uncertainty gating:** switch to slow, guarded moves when visual heightmaps are noisy; ask for human help after N retries.
- **Owner:** Manipulation lead.
- **Triggers/metrics:** Laundry F1 < 0.90 or retries > 3 → escalate; cable curvature violations > 0 per 100 cycles.
- **Playbook:** Pause → suggest fixture → downgrade speed/force → photo prompt to user → log hard-case for local adapter.

R2. Open-set safety (unknown humans/pets/objects)

- **Why it matters:** The world contains unseen classes; misclassification near people is unacceptable.
- **L/I:** L=High, I=High → **Residual:** L=Med, I=Med
- **Symptoms:** Late yielding, near-misses in corridors, hesitation loops.
- **Mitigations:**
    - **SSM hard caps** fused from vision+depth+capacitive; unknown classes inflate separation and reduce speed.
    - **Conformal prediction + ensembles** for calibrated abstention; tiered policy (slow → alternative plan → ask → refuse).
    - **Red-team drills** (mirrors, TV people, costumes, reflective knives) before each release.
- **Owner:** Perception safety lead.
- **Triggers/metrics:** Any proxemics breach in testing → hold release; abstention accuracy > 95% in OOD suite; zero unsafe contacts in pilots.

R3. Tool hazards (knives, heat, power tools)

- **Why it matters:** Small envelope errors have high downside (cuts, burns, property damage).
- **L/I:** L=Med, I=High → **Residual:** L=Low, I=Med
- **Symptoms:** Over-torque on jar lids, unlatched blade covers, energization without policy.
- **Mitigations:**
    - **Cryptographic tool interlocks** (identity → capabilities → envelope); rails off unless verified.
    - **Two-man rule** for sharp/heat classes; mandatory confirmations; **guarded trajectories** with tactile first-touch.
    - **Health-aware disable:** missing palm skin near blade disables knife skills.
- **Owner:** Safety kernel lead.
- **Triggers/metrics:** Interlock fault = stop + incident review; over-torque and heat limit violations = 0 in acceptance.

R4. Force runaway / contact overshoot

- **Why it matters:** Controller instability or sensor drift can exceed safe forces.
- **L/I:** L=Low, I=High → **Residual:** L=Low, I=Low
- **Mitigations:** Series elasticity + joint torque sensing; **1–3 ms reflex clamp** on over-envelope; monthly deep calibration; duplicate force channels for cross-check.
- **Owner:** Controls lead.
- **Metrics:** Peak overshoot < 10 N beyond envelope in kitchen suite; reflex latency median < 3 ms.

R5. Fall risk with payloads

- **Why it matters:** Dropping objects can injure or damage floors.
- **L/I:** L=Med, I=Med → **Residual:** L=Low, I=Low
- **Mitigations:** Push-recovery, widened stance under load, wheel-assist for long hauls; slip detectors in soles; payload-aware ZMP limits.
- **Owner:** Locomotion lead.
- **Metrics:** Mean fall rate < 1 per 200 h mixed-home; 0 drops in 100-cycle kitchen runs.

10.2 Supply chain, floor damage, voice spoofing, liability

R6. Supply-chain fragility (actuators, sensors, batteries)

- **Why it matters:** Single-source reducers, motors, or tactile stacks can halt production and raise costs.
- **L/I:** L=Med, I=High → **Residual:** L=Low, I=Med
- **Mitigations:**
    - **Dual-sourcing** critical parts; footprint-compatible alternates; long-lead buffers (90–120 days).
    - **Design for substitution:** common mountings, flexible bus voltages; reduce exotic materials.
    - **Supplier audits** for yield/quality; lifetime buys for at-risk components.
- **Owner:** Supply chain director.

- **KPIs:** Fill rate ≥ 95%; lead-time EWMA alarms; requal cycle < 60 days.

R7. Floor and property damage (scratches, spills, collisions)

- **Why it matters:** Home trust hinges on not marring floors or furnishings.
- **L/I:** L=Med, I=Med → **Residual:** L=Low, I=Low
- **Mitigations:**
    - **Compliant soles** with non-marking compound; low normal force during turns; wheel-assist torque caps.
    - **Spill field** in world model; slow zones on slippery surfaces; cloth-first cleanup policies.
    - **No-go geofences** for fragile areas (antique rugs, art zones).
- **Owner:** HRI + mobility.
- **Metrics:** 0 scuff incidents in acceptance; slip-related safe-stops logged and addressed via slower mode.

R8. Voice spoofing, mishearing, or coercion

- **Why it matters:** Adversaries (or kids) could issue unsafe commands.
- **L/I:** L=Med, I=High → **Residual:** L=Low, I=Med
- **Mitigations:**
    - **Wake-word + speaker verification** (on-device) for privileged actions; **small, robust safety lexicon**.
    - **Policy engine** binds tool enable/doors/heat to **owner/adult** roles; silent hours; rate limits.
    - **Gesture parity:** safety gestures override voice (stop palm, hands-to-head).
- **Owner:** HRI security lead.
- **Metrics:** Impersonation success < 1% in red-team; false accept of unsafe requests ≤ 0.1%.

R9. Privacy breach (data egress or over-collection)

- **Why it matters:** Loss of social license and legal exposure.
- **L/I:** L=Low, I=High → **Residual:** L=Low, I=Med

- **Mitigations:** Local-first processing; **120 s rolling black-box only**; encrypted receipts; opt-in federated with DP; on-device redaction of faces/screens.
- **Owner:** Data stewardship lead.
- **Metrics:** Zero unconsented egress in audits; DP budget (epsilon) published and enforced.

R10. Liability and insurance gaps

- **Why it matters:** Unclear fault lines can stall adoption and inflate costs.
- **L/I:** L=Med, I=High → **Residual:** L=Low, I=Med
- **Mitigations:**
    - **Signed safety cards** (domains, limits, invariants) + **audit receipts** per task.
    - Product liability coverage; recommended owner riders (equipment breakdown, cyber).
    - **Incident protocol:** safe-stop, black-box snapshot, human-readable summary, rapid vendor review.
- **Owner:** Legal & risk.
- **Metrics:** Claim rate per 10k operating hours; time-to-resolution; premium deltas tied to safety telemetry.

R11. Model/skill regressions after updates

- **Why it matters:** A “better” model may erode safety or a rare skill.
- **L/I:** L=Med, I=Med → **Residual:** L=Low, I=Low
- **Mitigations:** Canary channel, **shadow mode** comparison, **monotonic safety checks**, quick rollback; versioned **skill store signing**.
- **Owner:** ML release manager.
- **Metrics:** No safety regression vs prior; canary pass rate ≥ 99%; rollback MTTR < 1 h.

R12. Battery thermal or aging faults

- **Why it matters:** Fire risk and sudden brownouts.
- **L/I:** L=Low, I=High → **Residual:** L=Low, I=Med
- **Mitigations:** Cell-level fusing, BMS with SOH/SOC modeling, phase-change pads, pack isolation; **hot-swap** path; charge in supervised dock.

- **Owner:** Power systems.
- **Metrics:** 0 thermal events; SOH trend alarms; brownout-to-safe-stop success 100%.

R13. Ecosystem drift (appliances, layouts, standards)

- **Why it matters:** New device designs or building codes can break assumptions.
- **L/I:** L=Med, I=Med → **Residual:** L=Low, I=Low
- **Mitigations:** Appliance API partnerships; **affordance learning** over internet-scale videos; household commissioning that scans and stores fixture params.
- **Owner:** Partnerships + perception.
- **Metrics:** Commissioning success ≥ 99%; annual fixture re-scan prompts; adapter updates within 60 days of major appliance shifts.

Verification cadence

- **Quarterly risk review:** update L/I, add incidents, refresh mitigations.
- **Pre-release gate:** red-team passes R2/R8, tool interlock tests R3, privacy audit R9, regression checks R11.
- **Field telemetry:** dashboards show near-real-time breaches of budgets (falls, tele-assist minutes, abstentions, proxemics flags).

Bottom line: The risks that make open-world robots hard—**deformables, the unknown, and sharp/hot tools**—are managed by **invariants in the kernel**, **calibrated abstention**, and **operational fixtures and policies**. Non-technical risks—**supply, privacy, spoofing, liability**—are treated as first-class engineering problems with concrete owners, budgets, and auditable controls.

11. Validation Suite and Acceptance Tests

The validation suite converts requirements into **repeatable, scored trials** that a unit must pass before site deployment and after any safety-relevant update. Trials run on-instrumented courses, capture **traceable artifacts** (kernel traces, receipts, black-box snapshots), and produce **go/no-go** verdicts with margin. Unless stated, all metrics are 95%-confidence across ≥3 independent runs.

11.1 Mobility trials, manipulation cycles, social safety

11.1.1 Mobility trials

Courses

- **Mixed-home loop:** thresholds (15–25 mm), tight doors (≤760 mm), rugs, wet tile, low-pile carpet, clutter.
- **Verticals:** residential stairs (rise 180 ± 10 mm, run 280 ± 10 mm), handrail present; 12° ramps, curb cuts.
- **Transit fixtures:** elevator mock (door width 900 mm, dwell 6 s), hallway turns (90°/135°) with variable traffic.

Procedures

1. **Baseline traverse (45 min):** walk 0.8 m/s target; carry empty tray; random obstacle seeds each run.
2. **Load traverse (20 min):** two-hand 10 kg box; include ramps and elevator.
3. **Impulse + recovery:** lateral 80 N at pelvis (3 angles); forward 60 N while stepping.
4. **Surface adhesion:** wet tile and powder-dust patches; friction classifier must flip to “slow mode” pre-contact.
5. **Elevator ingress/egress:** 30 cycles with mixed passenger flow; doorway clearance ≤30 mm toe margin.

Pass criteria

- **Falls:** 0 per run; aggregate mean fall rate < 1/200 h.
- **Push-recovery:** no step loss in ≥95% of impulses; time-to-stabilize < 250 ms median.
- **Ingress/egress:** ≥97% first-try success; no door strikes.
- **Energy:** ≤600 W average over mixed loop (pack SoC and power trace supplied).
- **Kernel:** no SSM breach flags; braking distance margin ≥10% at all times.

Artifacts

- ESDF snapshots, footstep plans, kernel SSM tracepoints, IMU/force plates, receipts for load periods.

11.1.2 Manipulation cycles

Suites

- **Kitchen set:** unload/load dishwasher (12 plates, 8 glasses, 16 utensils), cabinet put-away, jar open/close, wipe counters.
- **Laundry set:** sort (colors/delicates), wash-to-dry handoff (simulated), fold (20-item mixed), put-away in labeled drawers.
- **Cables & fixtures:** thread 3 m cable through guides, plug insertion (Type A/C), lever/button panels with guarded force.
- **Kitting & totes:** pick-to-bin (50 items), label application, small-box tape close (with jig).

Procedures

1. **100-cycle stress run** per micro-task; randomized object poses and bin assignments.
2. **Deformable focus block:** cloth heightmap degraded (occlusions, crumples) at pre-set seeds.
3. **Tool class block:** knife proxy with sheath; heat proxy at 60 °C handle; policy must demand confirmations and apply low-force envelopes.

Pass criteria

- **Rigid grasps:** ≥98% success; ≤1.5 grasp attempts avg.
- **Glasses/fragile:** 0 drops over 100 cycles; contact overshoot < 10 N beyond envelope.
- **Laundry:** fold F1 ≥ 0.90; drawer placement ≥97% correct.
- **Cable/thread:** ≥95% success; no over-bend beyond spec.
- **Guarded insertions:** ≥97% without cross-thread; torque peaks within limits.
- **Knife/heat policies:** confirmations required; interlocks verified; 0 unauthorized energizations.

Artifacts

- Tactile summaries (200 Hz), force envelopes, slip events, policy hashes, per-task receipts (with IDL and outcomes).

11.1.3 Social safety & HRI

Scenarios

- **Corridor encounters:** single and multi-party crossings, pet proxies darting, stroller pass-by.
- **Doorway courtesy:** yield at thresholds, bidirectional traffic with tray-carry.
- **Personal space intrusions:** actors step into Z1/Z0 zones unexpectedly.
- **Accessibility:** ASL subset commands, large-print UI confirmations, silent-mode operation during “quiet hours”.

Procedures

1. **20-min flow run** with randomized human speeds (0.6–1.4 m/s), occlusions, and sudden stops.
2. **Pet proxy darts** (0.7–1.5 m/s) from blind corners; 30 events.
3. **Child mannequin** erratic path; robot must down-mode speed and increase separation.
4. **Gesture/e-stop battery:** stop palm, hands-to-head, voice e-stop with accents/noise profiles.

Pass criteria

- **Proxemics:** ≥99% compliance; zero unsafe contacts.
- **Yielding:** time-to-yield < 500 ms median; no corridor deadlocks > 3 s.
- **E-stop:** motors-idle < 0.8 s worst case; median < 0.5 s.
- **Accessibility parity:** task success within 2% of voice baseline.

Artifacts

- Kernel proximity states, SSM zone transitions, gesture/voice recognition logs, black-box snapshots for near-misses (should be none).

11.2 Household “new home in a day” and “errand loop”

11.2.1 “New home in a day” (commission → chores → tidy)

Purpose
Validate that a fresh unit can map, learn household conventions, and complete a meaningful day's work without code changes.

Environment
3-bed/2-bath apartment analog (≈90 m²): kitchen with dishwasher/oven, laundry closet, two bedrooms, mixed flooring, typical clutter. Owner provides default preferences (e.g., "mugs in cabinet 3").

Phase A — Commissioning (≤90 min)

- **Map & label:** scan rooms, identify major fixtures and storage.
- **Norms:** bind initial rules (cabinet/drawer assignments, quiet hours).
- **Acceptance:** mapping coverage ≥95%; object category recall ≥90% for kitchen; zero private-zone violations.

Phase B — Workday (≤6 h active)
Tasks issued as natural language; robot compiles to IDL.

- **Kitchen:** unload dishwasher → sort/put-away; wipe counters; take out trash.
- **Laundry:** collect baskets → wash/dry handoff (simulated) → fold → put-away.
- **Tidy:** living room reset (books, toys), vacuum handoff (start robot vac if available), water two plants (measured pour).
- **Ad-hoc:** user injects two repairs ("use cabinet 2 for mugs", "skip delicates").

Success metrics

- **Task success:** ≥97% overall across ≥40 subtasks.
- **Repairs:** ≥90% successful recovery on first correction; household norm updated (future uses pick cabinet 2).
- **Safety:** 0 unsafe contacts; a maximum **tele-assist ≤ 30 min total** (target ≤15 min).
- **Energy:** completes 6 h active with ≤1 hot-swap; docking without human nudge.

Artifacts

- Full receipts bundle (IDL plans, confirmations, outcomes), map hash before/after, privacy attestations, consumption of failure budgets (should be within limits).

11.2.2 “Errand loop” (elevator → sidewalk → pickup → return)

Purpose
Demonstrate constrained public navigation and commerce handoff under safety and policy controls.

Route
Apartment → elevator → lobby → sidewalk (100 m) → store counter pickup → elevator → apartment.

Preconditions

- Building and sidewalk mapped with safe corridors and **no-go zones**; store has API/QR pickup; policy forbids street crossing and running.

Procedures

1. **Elevator choreography:** wait, enter without blocking others, position to the side; 10 cycles at random floors.
2. **Sidewalk segment:** wheel-assist mode; keep right; yield to pedestrians, dogs, strollers; handle curb cut and construction barrier.
3. **Pickup:** present QR, confirm order verbally, secure small parcel in chest bin; verify with photo receipt.
4. **Return:** reverse route; hand parcel to owner with confirmation.

Pass criteria

- **Throughput:** round-trip ≤ 20 min median without rushing; ≤ 25 min p95.
- **Yielding & proxemics:** 0 contact events; proactive slowdowns on approach; no stops > 5 s without cause.
- **Transfer integrity:** parcel secured; no drops; photo receipt and store confirmation logged.
- **Policy adherence:** no entry into street; no use of disallowed shortcuts; silent mode in elevator if requested.
- **Public etiquette:** attention cues rated “clear” by ≥95% of test participants.

Artifacts

- Route plan and costmaps, SSM logs along sidewalk, QR/auth transcripts (redacted), parcel handling receipts, black-box snapshot triggers (none expected).

11.2.3 Scoring and gates

Scoring rubric (100 pts)

- Mobility (25): falls, ingress, energy margin.
- Manipulation (25): success rates, force envelopes, deformables.
- Social safety (20): proxemics, yielding, e-stop.
- Household day (20): completion, repairs, tele-assist minutes.
- Errand loop (10): etiquette, timing, policy adherence.

Go/No-Go

- Any **safety invariant breach** (SSM violation, tool interlock fault, unauthorized data egress) → **automatic NO-GO**.
- Otherwise, pass if **≥ 90 pts** and all minimums met; **85–89 pts** enters remediation (targeted fixes, limited redeploy); **< 85 pts** requires redesign.

Regression policy

- Post-update, run a **reduced cert**: top-risk mobility course, kitchen micro-suite, social battery, and both scenario smoke tests. Failures auto-rollback.

Summary: The validation suite is a contract: **if the robot can't pass these trials, it doesn't ship**. Mobility and manipulation benchmarks test physics and dexterity; social safety proves legibility and restraint; the **"new home in a day"** and **"errand loop"** scenarios verify real utility under policy and privacy constraints. The artifacts they produce let owners, integrators, and insurers verify performance without peeking into private life.

12. Deployment Roadmap (P0 to P3)

The roadmap builds capability and trust in **bounded, auditable steps**. Each phase has **go/no-go gates**, **owners**, **artifacts**, and **KPIs**. Regression at any gate rolls back to the last safe release.

12.1 Lab → Pilot Homes (P0 → P1)

P0 — Lab prototyping (0–12 months)

Objectives

- De-risk core stacks: hybrid locomotion, diffusion+TAMP manipulation, safety kernel v1, IDL + confirmations, tactile servoing, FRU swaps.

Environments

- Sandboxed kitchen/laundry bays; mixed-home mobility course; red-team arena (mirrors, knives/heat proxies, pet/child proxies).

Must-hit milestones

- Mobility: mean fall rate < 1/200 h on mixed course; elevator ingress ≥ 97%.
- Manipulation: kitchen micro-suite ≥ 97% success; laundry F1 ≥ 0.90.
- Kernel: SSM, force ceilings, interlocks verified; e-stop ≤ 0.8 s p99.
- Learning: foundation model v1; on-device adapters gated by mini cert.

Artifacts

- Safety Card v0, Model/TrainRun Cards, risk register R1–R5 closed to residual L≤Med.
- Pilot bill of materials (BOM) + FRU catalog; service playbooks v0.

Exit gate

- Pass 11.1 core trials and 11.2 smoke tests in the lab; privacy audit passes; incident rate = 0 in acceptance.

P1 — Pilot homes (12–24 months)

Objectives

- Prove **“new home in a day”** workflow, low tele-assist, and owner trust signals.

Cohort

- 20–40 screened households (varied layouts, pets/children, appliances); NDA + IRB-style consent.

Operations

- White-glove install; commissioning ≤ 90 min; weekly check-ins; Tier 0/1 service SLAs.
- Optional **skill channels**: stable (kitchen), beta (laundry), canary (tidy).

KPIs (rolling 4-week)

- Task success ≥ 97%; tele-assist ≤ 30 min/100 h (target 15); availability ≥ 99.0%.
- Safety: 0 unsafe contacts; refusal accuracy ≥ 99% for unsafe asks.
- Trust: NPS-like ≥ +40; "clear behavior" ≥ 95% in HRI surveys.

Gates

- 90% of homes pass 11.2 "day-in-life" without staff on-site.
- Privacy posture unchanged (no new data egress); receipts accepted by two external reviewers (insurer + academic partner).

Rollback triggers

- Any SSM breach, tool-interlock fault, or privacy incident → immediate pause, black-box review, hot-fix or revert.

12.2 Back-of-House Services (P2)

P2 — Semi-structured commercial pilots (24–36 months)

Targets

- Hotels (housekeeping carts/linen rooms), clinics (supply runs), big-box stores (off-hours restock), office facilities.

Why now

- Repetitive tasks, controllable layouts, clear ROI, existing safety doctrine.

Site readiness checklist

- **Workcells** with geofences and signage, low-traffic windows, pallet/tote specs, elevator/door APIs where possible.
- **Policy pack**: actor roles, hazard tools disabled by default, silent hours, audit recipients.

Pilot structure

- 5–8 sites, 2–6 units/site, 3–6 months each; mixed vendor fixture kits (carts, guides).
- Tele-ops desk with 1:8 operator:robot ratio (target 1:12 by end of P2).

KPIs

- Throughput: ≥ 85% of human baseline first month → ≥ 110% by month three (for targeted tasks).
- MTBI ≥ 8 h; availability ≥ 99.5%; downtime ≤ 10 h/year (annualized).
- Safety: 0 OSHA-recordable events; near-misses captured with black-box and corrective actions logged within 72 h.

Economic yardsticks

- Cost/hour ≤ $8–$12 all-in at 50% utilization; breakeven < 24 months for restock/kitting cells.

Compliance & partners

- Align with site EHS; ANSI/ISO robot safety review; insurer sandbox.
- Integration owners per site (facilities + IT) with weekly JSA (Job Safety Analysis) updates.

Exit gate

- Two sites renew/expand; three third-party attestations (EHS, insurer, standards body) accept receipts in lieu of raw video; red-team OOD suite passes with abstention ≥ 95%.

12.3 Broad Public Deployments & Certification (P3)

P3 — Scale-out and certification (36–60 months)

Objectives

- Transition from pilots to **repeatable deployments** in homes and services with cataloged workcells and auditable compliance.

Scale levers

- **SKU-ized deployments:** kitchen pack, laundry pack, restock pack; each with fixed FRUs, fixtures, acceptance scripts.

- **Skill store governance:** stable/beta/canary lanes; monotonic-safety CI; fleetwide rollback in < 1 h.
- **Partner ecosystem:** appliance APIs, building/elevator vendors, facilities integrators.

Certification & assurance

- Third-party **Safety Case**: machine-readable claims tied to Validation IDs and Kernel/Policy hashes.
- Self-declaration to ISO/ANSI with notified-body reviews where applicable; publish Safety Card updates.
- Data protection audits (privacy-by-default) and DP budgets for federated learning disclosed.

Public-space constraints

- Constrained **errand loops** only (no traffic); municipal MOUs for sidewalks; etiquette studies with ≥ 95% "clear behavior" rating.
- Incident playbook co-written with insurers; public hotline + OTA hot-fix path.

SLOs at scale

- Home: availability ≥ 99.0%, annual downtime ≤ 20 h, tele-assist ≤ 30 min/100 h.
- Services: availability ≥ 99.5%, downtime ≤ 10 h/year, operator ratio ≤ 1:12.
- Fleet: kernel availability ≥ 99.99%; zero critical interlock faults; revocation MTTR < 24 h.

Org & process

- **Release board** (Safety, ML, HRI, Legal) signs off weekly; any safety-relevant change requires canary + reduced cert pass.
- **Field reliability** team runs EWMA alarms (falls, refusals, abstentions, MTBI) and schedules proactive FRU swaps.

Exit criteria for P3 (sustained run-rate)

- ≥ 1,000 deployed units with 90-day rolling compliance to SLOs;
- Claims rate below insurer threshold; premiums stable or declining;
- Independent lab reproduces key acceptance tests on random retail units.

Cross-phase Enablers (apply P0–P3)

- **Telemetry discipline:** receipts, kernel traces, health metrics—never raw household video by default.
- **Hot-swap logistics:** regional depots, 72-hour arm/hand/battery availability.
- **Comms & trust:** transparent changelogs, Safety Card diffs, privacy posture unchanged unless owner opts in.
- **Kill-switch posture:** remote revocation for defective skills/models; owner-visible and auditable.

Summary: P0 proves physics and safety in the lab; P1 earns household trust; P2 demonstrates ROI in semi-structured services; P3 codifies certification, partners, and scale. Each step raises the bar while keeping **safety invariants, policy control, and privacy** intact.

13. Ethics, Law, and Standards

Ethics, law, and standards are baked into the system as **requirements**, not afterthoughts. This section operationalizes consent and privacy, maps our controls to relevant safety norms, and defines insurance/incident workflows that protect people while enabling transparent, auditable operations.

13.1 Consent, privacy, and data minimization

Consent-by-design

- **First-run contract:** plain-language walkthrough of capabilities, limits, and what the robot does *not* do (e.g., no default recording, no face-based policing). Separate, granular toggles for: diagnostics, hard-case caches, federated learning, and third-party audit sharing.
- **Household roles:** owner / adult / guest / child, each with default permissions. Child mode defaults to **no tool enable**, **no data sharing**, and **extra confirmations**.
- **Just-in-time prompts:** any new data use (e.g., enabling “hard-case cache”) triggers a one-time modal, not buried in settings.

Data minimization & purpose limitation

- Default **local-first** processing; no persistent video/audio capture. Rolling black-box buffer (≈120 s) exists solely for safety forensics.
- Receipts log *actions* (who/what/when/where/force bands), not raw sensor content.
- Optional cloud/fleet features are **opt-in** with explicit purpose, retention, and revocation paths.

Transparency & control

- **Privacy dashboard:** shows what's on, where data lives, retention timers, and who can request it. One-tap: pause learning, purge caches, export/corroborate receipts.
- **Right to be forgotten:** purges local adapters, hard-case caches, and cryptographically tombstones federated contributions in future aggregates.

Data retention

- Receipts: 12 months by default (owner-tunable); kernel/error traces aligned to warranty/service windows; black-box snapshots retained **only** on incidents or owner request.
- No silent expansion of scope: changes to retention or telemetry require explicit owner consent and versioned policy diffs.

Children, guests, and bystanders

- **Bystander-safe defaults:** non-identifying person tracking only; no bystander profiles.
- **Guest code** (QR or phrase) grants temporary interaction without enrolling accounts; ends automatically after visit.

Security posture

- Secure boot, signed skills/adapters, hardware-backed keys, encrypted storage and transport.
- Principle of least privilege for on-device services; air-gap mode available (no radios) with degraded functionality banner.

Ethical guardrails

- Hard denials: weapons handling, surveillance resale, person-specific pursuit/policing, coercive uses.

- Refusal rationales are explained in natural language and attached to receipts.

13.2 Alignment with ISO/ANSI safety norms

We implement to the spirit and letter of current collaborative/service robot standards; specific citations below guide design and verification. (Exact editions will be listed in the Safety Case.)

Core machinery & robot safety

- **ISO 12100** (risk assessment): Used for hazard identification → risk evaluation → risk reduction; outputs tie to kernel invariants and validation suites.
- **ISO 10218-1/-2** / **ANSI/RIA R15.06** (industrial robots & integration): Inform safe states, emergency stops, interlocks, and integrator responsibilities for semi-guarded cells.
- **ISO/TS 15066** (collaborative robot guidance): Basis for speed-and-separation monitoring (SSM), quasi-static/ transient contact limits, and pain-threshold maps that set our force ceilings.
- **ANSI/RIA R15.08** (mobile robots): Governs navigation in shared spaces, braking distances, and traffic management for the wheel-assist base; we document d_stop margins and SSM zoning accordingly.
- **ISO 13482** (personal care/service robots): Applies to domestic/assistive contexts—defines limits for contact, stability, and protective measures around non-expert users.

Functional, electrical, and software safety

- **IEC 61508 / IEC 62061 / ISO 13849-1**: Our safety kernel targets a documented Performance Level (e.g., PL d, Cat. 3) for e-stop, tool rails, and speed limiting; dual-channel sensing and diagnostics justify the claimed PL.
- **IEC 60204-1** (electrical machinery safety): Drives enclosure, wiring, and protective bonding; tool rails isolated and interlocked.
- **EMC/EMI** (IEC 61000 series; FCC/RED): Ensures immunity/emissions won't degrade safety sensing.
- **Battery & charging** (UN 38.3 transport, UL/IEC cell/pack standards): Pack/BMS design, hot-swap, and dock isolation tested against thermal and fault scenarios.

Human factors and usability

- **ISO 9241 / IEC 62366-inspired** usability processes: We run IRB-style HRI studies for clarity of cues, proxemics, and accessibility parity; results feed risk reduction and labeling.

Safety Case & conformity

- **Machine-readable Safety Case** links each norm to: design controls (kernel/policy features), verification artifacts (11.x trials), and traceability (model/policy hashes).
- **Change control:** Any update that alters a safety claim triggers a reduced re-cert suite and Safety Card diff shipped to owners.

13.3 Insurance and incident response

Insurance stack (recommended)

- **Product liability** (vendor-held): covers design/manufacture defects.
- **General liability** (operator/venue): premises and third-party property.
- **Equipment breakdown**: actuators/compute/battery failures.
- **Cyber**: data breach, unauthorized control, incident response.
- **Workers' comp interface** (back-of-house): clarifies coverage when robots share workcells.

Telemetry-based assurance

- Insurers receive **receipts and safety metrics** (not raw sensor data): kernel availability, SSM events (none), e-stop latencies, MTBI/MTTR, force-envelope violations (none), and red-team pass rates. Usage-based pricing can reward demonstrated safety.

Incident taxonomy & SLAs

- **Severity 1 (S1):** injury requiring medical attention, fire, uncontrolled tool event.
    - Immediate safe-stop, secure black-box snapshot, 24/7 hotline ping; vendor engages within **1 hour**; report package delivered in **72 hours** with corrective actions.
- **Severity 2 (S2):** property damage > owner threshold, near-miss with kernel intervention.
    - Safe-stop, snapshot, on-site review in **1 business day**, fix in **5 business days**.
- **Severity 3 (S3):** nuisance failure (task aborts, minor scuffs).

    - Logged, triaged in the next software hotfix cycle.

Forensics & disclosure

- **Black-box and receipts** are signed and encrypted; owner holds keys. Vendor can only decrypt with owner consent or lawful process. Third-party verifier tool reproduces motion envelopes and decisions without exposing private video.
- **Regulatory reporting**: pre-authored templates for jurisdictions requiring incident notification (e.g., consumer safety agencies, workplace safety authorities).

Corrective and preventive action (CAPA)

- Root-cause analysis uses kernel traces, model/policy hashes, and FRU health.
- **Global mitigations** (e.g., tighter knife envelopes, adapter rollback) roll out via signed updates with release notes and Safety Card deltas.
- **Revocation path:** defective skills/models placed on a signed CRL; fleet revocation MTTR target **< 24 h**.

Contractual & site governance

- **Site Safety Plan**: roles, quiet hours, tele-assist rules, geofences, housekeeping for slip/spill risks.
- **Allocation of responsibility**: who owns fixtures, who signs policy changes, who receives incident notices.
- **De-installation & end-of-life**: certified data purge, eco-disposal/recycling, and insurance closure letter.

Ethical disclosure & community trust

- Public **Safety Cards** summarize domains, limits, known failure modes, and incident statistics (aggregated).
- Clear channel for vulnerability reporting (security.txt equivalent) and bug bounties on safety-relevant issues.

Summary: Ethical use is enforced by **consent-first UX and strict data minimization**; legal diligence comes from **traceable alignment to ISO/ANSI norms** with machine-readable Safety Cases; and resilience is ensured through **insurance-aligned telemetry and rigorous incident**

response. Together, these make the robot not just capable, but **accountable and trustworthy** in homes and public spaces.

14. Related Work and Differentiation

This section situates our approach within current humanoid and service-robot efforts across industry and academia, and explains why a **safety-first, audit-ready** architecture is the decisive lever for real-world adoption—not merely a compliance layer.

14.1 Humanoid trends and gaps

Trends across the field

- **Form factor convergence:** Adult-sized bipeds with 5-finger hands and human-reach envelopes to exploit existing tools, doors, and appliances.
- **Learning-centric stacks:** Shift from hand-coded pipelines to **foundation visuomotor models**, imitation learning from human/robot video, and policy fine-tuning from limited tele-op data.
- **Factory-first deployment:** Most vendors target **internal or partner factories** (kitting, line-feeding, packout) before household or public venues—structured layouts and clearer ROI.
- **Tactile renaissance:** Renewed investment in **dense fingertip and palm sensing**, slip detection, and whole-body compliance for safe contact-rich manipulation.
- **HRI formalization:** Proxemics, gaze cues, and "legible motion" are moving from research to product requirements; voice + gesture hybrids replace voice-only.
- **Safety kernel revival:** Speed-and-separation monitoring (SSM), torque ceilings, and certified e-stop paths are re-emerging as first-class runtime systems rather than peripheral PLCs.
- **Model lifecycle tooling:** Early ecosystems for **skill stores**, model versioning, and fleet updates are appearing, though governance remains uneven.

Common capability gaps

- **Deformables at scale:** Laundry, cables, bags remain brittle; few publish long-horizon, high-success cloth pipelines outside controlled demos.

- **Open-set robustness:** Many systems lack calibrated abstention and clear fallback policies when encountering unknown objects, mirrors, TV people, or unusual lighting.
- **Safety invariants vs. ML autonomy:** Learned policies often bypass or weaken physics-level guards; consistent **millisecond reflex clamping** and kernel-mediated envelopes are rare.
- **Tool governance:** Knife/heat/power tool use is typically demoed without cryptographic interlocks, capability declarations, or role-aware policies.
- **Auditability & privacy:** Telemetry is either too sparse for insurers/regulators or too invasive for homes; few solutions offer **receipts** (action-level logs) without raw video.
- **Serviceability:** Distal-mass reduction, FRUs, and hot-swaps are under-specified, inflating downtime and TCO for pilots.
- **Certification narrative:** Mapping features to ISO/ANSI claims is ad hoc; few provide **machine-readable safety cases** or Safety Card deltas per release.

Academic baselines vs. productization

- Labs lead on **dexterous manipulation, diffusion policies, tactile servoing**, and **foundation models**. Product teams must fuse these into **auditable stacks** (policy kernels, receipts, interlocks) and withstand multi-hour duty cycles, thermal limits, and wear—all historically outside academic scope.

14.2 Why a safety-first, audit-ready stack is decisive

1) Converts capability into deployability

- A robot that succeeds 98% of the time but can violate force envelopes or SSM once is **non-deployable**. Kernel-enforced **invariants** (speed, separation, torque, tool rails) make learned competence usable in mixed company—children, pets, the public.

2) Lowers insurance friction and speeds procurement

- **Receipts** (who/what/when/force bands/policy hash) plus a machine-readable **Safety Case** let insurers price risk without ingesting private video. Facilities teams can approve pilots with clear acceptance tests and post-incident forensics—no bespoke paperwork each time.

3) Enables privacy-by-default adoption in homes

- Local-first compute, **120-s black-box buffers only**, and per-role policy gates (owner/adult/guest/child) resolve the core household objection: "I don't want a camera streaming my life." Audit without surveillance is the social license.

4) Scales learning without eroding safety

- On-device adapters + **federated learning with DP** grow capability, while **shadow mode**, **monotonic-safety checks**, and canary channels prevent regressions from silently slipping into production. Most stacks lack this coupling between ML lifecycle and safety guarantees.

5) Makes tool use governable—not just possible

- **Cryptographic tool interlocks**, capability declarations, and role/time/place policies transform knives, heat, and power tools from PR demos into **bounded, attestable behaviors**. This is the difference between a prototype and a product that can pass an EHS review.

6) Drives down TCO and downtime

- Reliability is not only MTBF—it's **MTBI (intervention hours)** and **MTTR**. FRUs (hands/arms/soles/skin tiles), auto-cal, and health-aware derating keep units earning. Safety invariants reduce incident-induced downtime and claims.

7) Future-proofs certification and public deployment

- Standards evolve; cities demand etiquette and incident reporting; partners need integration attestations. A **policy- and receipt-native** design means adding a municipal rule, an elevator API constraint, or a new refusal policy is a **policy update**, not a re-architecture.

8) Aligns incentives across stakeholders

- Owners get control (policies, zones, roles). Operators get predictable service levels. Insurers get auditable metrics. Regulators get Safety Cases and incident hooks. ML researchers get a stable substrate where new policies can ship **without** risking invariants.

Our differentiation (summary)

- **Kernel-first runtime** with millisecond reflexes and SSM as hard constraints, not hints.
- **Policy engine** that binds tasks to actors, places, time, and tools—with cryptographic signing and rollback.

- **Receipts + black-box** for verifiable accountability without persistent recording.
- **Learning pipeline** that couples adapters/federation to safety gates (shadow mode, monotonic checks, red-team SLIs).
- **Serviceability & cost** engineered in (low distal mass, FRUs, consumables, depots) to hit sub-$16k/year operating targets in services.
- **Standards alignment** captured in a **machine-readable Safety Case**, updated with each release.

Bottom line: In a crowded field chasing dexterity and demos, the winning edge is not just a better hand or gait—it's a stack that **cannot quietly become unsafe**, can **prove** what it did, and can be **governed** by owners, insurers, and regulators. That is the difference between impressive videos and widespread, trusted deployment.

15. Conclusion: From Demos to Duty Cycles

15.1 What success looks like

Success is not another viral demo; it is a robot that **shows up, does work, leaves an audit trail, and earns trust**—day after day. Concretely, success means:

- **Duty-cycle reliability:** ≥97% task success across kitchen, laundry, kitting, and errand scenarios; MTBI ≥8 hours; availability ≥99% in homes and ≥99.5% in back-of-house.
- **Zero-surprise safety:** No unsafe contacts; e-stop ≤0.8 s worst-case; tool interlocks at 100% compliance; speed-and-separation never breached. Safety is *provably* monotonic across updates.
- **Legible behavior:** Proxemics compliance ≥99%, clear attention cues ≥95% human ratings, and corrections integrated on the first try ≥90% of the time.
- **Privacy by default:** Useful receipts without persistent video/audio. Owners control policies, roles, and zones; auditors verify outcomes without seeing private life.
- **Economics that scale:** Sub-$16k/year operating cost in services with FRUs and depots; three-year breakeven in homes or semi-structured sites at 2–5 hours/day utilization.
- **Governable evolution:** Skills ship through a signed store; adapters/federation improve capability under shadow-mode checks and red-team SLIs; a machine-readable Safety Case ties claims to validation IDs.

In short: **from spectacle to service**—the platform graduates when adding one more robot to the fleet predictably improves throughput and never widens the risk envelope.

15.2 Next experiments and community benchmarks

To accelerate the field and keep ourselves honest, we propose a compact set of **public, repeatable benchmarks** and **near-term experiments** aligned with this paper's invariants.

A. Kitchen & Laundry Core (KLC) Benchmark

- *Tasks:* dishwasher unload→put-away; jar open/close (torque-limited); laundry fold (20-item mixed)→drawer put-away.
- *Metrics:* success rate, force-overshoot violations (target 0), fold F1, slip events, cycle energy (Wh/task).
- *Artifacts:* signed receipts + kernel traces; no raw video required.

B. Deformables & Cables (DC-95)

- *Tasks:* cable routing with bend-radius caps; cloth flatten→fold with occlusions.
- *Metrics:* ≥95% success; zero over-bend; retries ≤3.
- *Focus:* tactile servoing and uncertainty-aware slow modes.

C. Social Safety Corridor (SSC)

- *Scenarios:* bidirectional traffic with pet/child proxies, stroller pass-by, doorway courtesy.
- *Metrics:* proxemics ≥99%, time-to-yield <500 ms median, e-stop ≤0.8 s p99.
- *Deliverable:* proxemics state machine trace + SSM logs.

D. New Home in a Day (NH1D)

- *Protocol:* cold start → map → chores (≥40 subtasks) → two mid-task repairs.
- *Metrics:* ≥97% completion, ≤30 min tele-assist/100 h, zero unsafe contacts.
- *Outcome:* proves utility without engineering on-site.

E. Errand Loop (EL-Public)

- *Route:* elevator→sidewalk (100 m)→counter pickup→return, no street crossing.
- *Metrics:* etiquette rating ≥95%, parcel integrity 100%, policy adherence 100%.

F. Monotonic-Safety Update Test (MSUT)

- *Method:* run KLC + SSC before/after model/skill updates in shadow mode.
- *Gate:* no degradation in refusal accuracy, e-stop latency, or force envelopes; else auto-rollback.

G. Tool Governance Drill (TGD)

- *Setup:* knife/heat/power tool with missing IDs, bad contexts, or child-present flag.
- *Metrics:* 100% denials or down-modes; interlock energization faults = 0.

H. Privacy & Audit Challenge (PAC)

- *Task:* third-party verifies performance using receipts/black-box only.
- *Success:* independent reproduction of outcomes without personal data exposure.

Near-term experiments

- **Shadow-mode federation:** run federated adapters alongside production for KLC; measure improvement vs. any safety trade-off.
- **Fixture-assist ablations:** quantify ROI of simple jigs (fold boards, cable guides) on DC-95 success and cycle time.
- **Cross-vendor reproducibility:** exchange NH1D homes between labs with only maps and policies; compare pass rates and tele-assist minutes.

By centering evaluation on **duty cycles, invariants, and auditability**, these benchmarks reward systems that are not just clever—but **deployable**. The community that adopts (and iterates) on this suite will turn open-world humanoids from promising prototypes into **trusted, working appliances** in homes and public spaces.

References

Standards & Safety Guidance

1. ISO/TS 15066:2016 — *Robots and robotic devices — Collaborative robots*. International Organization for Standardization. Automate Show+1
2. ISO 10218-1:2025 — *Robotics — Safety requirements — Part 1: Robots*. International Organization for Standardization. ISO+2ISO+2
3. ANSI/A3 R15.06-2025 — *Industrial Robots and Robot Systems—Safety Requirements*. Association for Advancing Automation (A3). The ANSI Blog+1
4. ANSI/RIA R15.08-1:2020 — *Industrial Mobile Robots—Safety Requirements—Part 1*. American National Standards Institute / RIA. ANSI Webstore+2AGV Network+2
5. ISO 13482:2014 — *Robots and robotic devices — Safety requirements for personal care robots*. International Organization for Standardization. ISO+1
6. IEC 61508 (Overview) — *Functional safety of E/E/PE safety-related systems*. International Electrotechnical Commission. IEC
7. ISO 13849-1:2015 — *Safety of machinery — Safety-related parts of control systems — Part 1*. International Organization for Standardization; overview resources. ISO+2Pilz+2
8. IEC 60204-1:2016 — *Safety of machinery — Electrical equipment of machines — Part 1*. International Electrotechnical Commission; overview resources. IEC Webstore+1
9. Marvel, J. A., *Implementing Speed and Separation Monitoring in Collaborative Robot Systems*, NIST (2017). NIST
10. OSHA Robotics Standards Overview (references to R15.06, etc.). U.S. Occupational Safety & Health Administration. OSHA

Visuomotor Learning, Foundation Models, & Datasets

11. Chi, C. *et al.* "Diffusion Policy: Visuomotor Policy Learning via Action Diffusion." *arXiv:2303.04137* (2023). arXiv+2arXiv+2
12. O'Neill, A. *et al.* "Open X-Embodiment: Robotic Learning Datasets and RT-X Models." *arXiv:2310.08864* (2023). arXiv+2arXiv+2
13. Brohan, A. *et al.* "RT-2: Vision-Language-Action Models Transfer Web Knowledge to Robotic Control." *CoRL 2023 / PMLR* (2023). arXiv+2Proceedings of Machine Learning Research+2

14. Google DeepMind. “Gemini Robotics and Gemini Robotics-ER for real-world robot tasks.” (news coverage, 2025). The Verge+1

Task & Motion Planning (TAMP) and Hybrid Planning

15. Kaelbling, L. P., & Lozano-Pérez, T. “Integrated Task and Motion Planning in Belief Space.” *IJRR* 32(9–10) (2013). Learning and Intelligent Systems Group+1

16. Garrett, C. R. *et al.* “Integrated Task and Motion Planning.” *Annual Review* / *arXiv:2010.01083* (2020/2021). arXiv+1

17. Toussaint, M. “Logic-Geometric Programming: An Optimization-Based Approach to Combined Task and Motion Planning.” *IJCAI* (2015). IJCAI

18. Phiquepal, C., & Toussaint, M. “Combined Task and Motion Planning under Partial Observability.” *ICRA* (2019). Argmin

19. (Survey) “A Survey of Optimization-based Task and Motion Planning.” *arXiv:2404.02817* (2024). arXiv

Uncertainty, Conformal Prediction, & Safety-Aware Control

20. Zhou, H., Zhang, Y., & Luo, W. “Safety-Critical Control with Uncertainty Quantification using Adaptive Conformal Prediction.” *arXiv:2407.03569* (2024). arXiv

21. Conformal-planning examples & summaries (NeurIPS 2023 poster). NeurIPS

Perception: Event & Multimodal Sensing

22. “Recent Event Camera Innovations: A Survey.” *arXiv:2408.13627* (2024). arXiv

23. IEEE-RAS T-RO Special Issue: *Event-Based Vision for Robotics* (call/overview, 2025). IEEE Robotics and Automation Society

Tactile Sensing & Robotic Skin

24. Yuan, W. *et al.* “GelSight: High-Resolution Robot Tactile Sensors for Dexterous Manipulation.” *Sensors* 17(12):2762 (2017). MDPI+1

25. Mandil, W. *et al.* “Tactile-Sensing Technologies: Trends, Challenges and Outlook.” *Sensors* (2023). PMC

26. Li, R. *et al.* “Implementing Monocular Visual-Tactile Sensors for Robust Manipulation.” *Sensors* (2022). PMC

27. (Overview) DIGIT tactile sensor (GelSight/Meta AI) — product note. pages.gelsight.com

28. Gomes, D. F. *et al.* “Beyond Flat GelSight Sensors: Simulation of Optical Tactile Sensors on Curved Surfaces.” *RSS* (Proc., 2023). Robotics Proceedings

HRI, Proxemics, & Social Navigation

29. Dragan, A. D., Lee, K. C. T., & Srinivasa, S. “Legibility and Predictability of Robot Motion.” *HRI* (2013); extended resources. Robotics Institute CMU+2ACM Digital Library+2

30. Singamaneni, P. T. *et al.* “A Survey on Socially Aware Robot Navigation: Taxonomy and Future Challenges.” *arXiv:2311.06922* (2023). arXiv

31. Gao, Y. *et al.* “Evaluation of Socially-Aware Robot Navigation.” *Sensors* (2022). PMC

Note: Where standards are paywalled, we cite official overview pages and trusted explanatory resources. For research papers, we link to publisher/arXiv or authoritative project pages.

Appendices

A. Interface Specifications (APIs & I/O)

A.1 Robot–App API (local, authenticated)

- **Transport:** mTLS over WebSocket (localhost proxy or LAN), JSON-RPC 2.0.
- **Auth:** Device-bound cert + per-user OAuth; roles: owner|adult|guest|child|tech.
- **Rate limits:** 20 RPS/user; 2 MB/min streaming receipts.
- **Key methods**
    - Plan.create(intent_text, options) -> plan_id
        - options: {safety_mode: "conservative|normal", timebox_min: int, privacy: {record:false}}
    - Plan.commit(plan_id) -> run_id
    - Run.pause(run_id), Run.cancel(run_id), Run.status(run_id)
    - Policy.get() -> policy_doc, Policy.update(policy_doc_signed)
    - Receipt.fetch(range) -> [receipt]
    - Map.snapshot() -> {hash, rooms, fixtures}
    - Diagnostics.bundle() -> ServiceBundle (.zip)
- **Events (server→client)**
    - need_confirmation(run_id, summary, risk_flags)
    - repair_request(run_id, question, choices)
    - hazard_detected(kind, location, action_taken)
    - run_completed(run_id, success, receipts_ref)

A.2 Tool Interlock Bus

- **Physical:** 8-pin blind-mate (12–48 V DC, I²C, 2× GPIO e-stop loop, tool-ID 1-Wire).
- **Handshake (≤200 ms):** READ_ID → CAPS (torque/temp/sharp edges flags) → ARM? (kernel policy check) → ARMED.
- **Safety:** Dual e-stop line continuity; ID mismatch or policy deny keeps rails de-energized.

A.3 Skill Package Manifest (signed)

```
{
  "skill_id": "com.example.kitchen.unload_dishwasher.v3",
  "kernel_abi": "1.9.2",
  "policy_min": "2025-08-01",
  "hashes": {"code": "sha256:...", "assets": "sha256:..."},
  "resources": ["perception/affordances.ron", "policies/knife_lowforce.json"],
  "channels": ["stable","beta","canary"],
  "safety_claims": ["SSM_enforced","force_ceiling_verified"],
  "signatures": {"vendor": "ed25519:..."}
}
```

B. Instruction DSL (IDL) Grammar & Examples

B.1 Grammar (EBNF)

```
Program     := Stmt | SEQ "(" Stmt { "," Stmt } ")"
Stmt        := TASK "(" Ident [ "," ArgList ] ")"
             | COND "(" Pred "," Stmt "," Stmt ")"
             | LOOP "(" Stmt "," Until ")"
             | PAR "(" Stmt { "," Stmt } ")"
ArgList     := Arg { "," Arg }
Arg         := OBJ "(" kvs ")"
             | LOC "(" kvs ")"
             | WHEN "(" kvs ")"
             | USE "(" kvs ")"
             | FILTER "(" Expr ")"
             | BOUNDS "(" kvs ")"     // force, speed, temp
```

```
          | DONE "(" kvs ")"
kvs       := Pair { "," Pair }
Pair      := Ident "=" Value
Value     := Number | String | Bool | Ident
Pred      := Expr
Expr      := Term { ( "AND" | "OR" ) Term }
Term      := Atom | "NOT" Atom | "(" Expr ")"
Atom      := Ident RelOp Value | InTest | Exists
RelOp     := "=" | "!=" | "<" | ">" | "<=" | ">="
InTest    := Ident "IN" "{" Value { "," Value } "}"
Exists    := "EXISTS(" Ident ")"
Until     := "until=" Pred | "iterations=" Number
Ident     := /[a-zA-Z_][a-zA-Z0-9_]*/
String    := quoted UTF-8
Number    := int | float
Bool      := "true" | "false"
```

B.2 Canonical Examples

Unload dishwasher then put mugs in cabinet 3, conservative force:

```
SEQ(
  TASK(unload_dishwasher, LOC(room=kitchen), BOUNDS(force_max_n=10)),
  TASK(put_away, OBJ(class=mug), LOC(region=cabinet_3), BOUNDS(force_max_n=8))
)
```

Only top rack; skip delicates; stop if guest arrives:

```
SEQ(
  TASK(load_dishwasher, FILTER(rack="top")),
  TASK(fold_laundry, FILTER(fabric!="delicate")),
```

```
  COND(EXISTS(guest_present)=true, TASK(pause), TASK(resume))
)
```

Parallel tidy with timebox and success check:

```
PAR(
  TASK(tidy_room, LOC(room=living_room), WHEN(deadline=20m), DONE(check=photo)),
  TASK(wipe_counters, LOC(room=kitchen), BOUNDS(speed="slow"))
)
```

B.3 IDL ↔ Receipt Binding (snippet)

```
{
  "receipt_id": "r_01F..",
  "plan_hash": "sha256:..",
  "idl_stmt": "TASK(put_away, OBJ(class=mug), LOC(region=cabinet_3))",
  "bindings": {"mug": ["obj#1342","obj#1347"], "cabinet_3": "loc#A2"},
  "envelopes": {"force_max_n": 8, "speed": "slow"},
  "outcome": "success",
  "started_at": "2025-10-24T15:03:11-05:00",
  "ended_at": "2025-10-24T15:05:02-05:00",
  "policy_hash": "sha256:..",
  "kernel_hash": "sha256:.."
}
```

C. Simulator Assets & Bench Setup

C.1 World Assets (URDF/SDFormat + semantic layers)

- **Rooms:** kitchen (dishwasher, cabinets, counters), laundry (washer/dryer geometry), beds/closets, corridors, elevator mock, lobby.
- **Objects:** plates, mugs (sizes), glassware, cutlery, jars with torque curves, towels, shirts, cables (Bézier rod models), totes, boxes.

- **Sensors:** RGB-D, fisheye, event cam plugins; tactile skin synthetic (taxel grids with noise models); foot force plates.

C.2 Physics & Contacts

- **Engine:** MuJoCo or Isaac Gym; soft contacts with Hunt–Crossley; friction cones with anisotropy for hardwood vs. tile.
- **Deformables:** Position-based dynamics for cloth; Cosserat rod model for cables with min-radius constraints.
- **Thermals (proxy):** tagged hot regions (stove handles @ 60 °C) to test avoidance.

C.3 Benchmark Scenarios (ready scripts)

- kitchen_unload_load_100cycles.scene
- laundry_fold_mixed_20items.scene
- cable_thread_guides.scene
- corridor_social_flow.scene
- elevator_ingress_multiparty.scene
- errand_loop_sidewalk_100m.scene

C.4 Data Export

- **Logs:** kernel tracepoints, SSM zones, torque/force, tactile summaries (200 Hz), success/failure tags.
- **Repro seeds:** fixed RNG for object poses, cloth crumple states, human agent paths.

D. Bill of Materials (indicative v1 BOM)

Costs are ballpark for small-batch (100–500 units); mass production targets lower by 20–35%.

Subsystem	Example Parts	Qty	Est. Unit Cost (USD)	Notes
Compute carrier	Edge AI SoC + GPU module, carrier board, ECC RAM	1	$1,800	250–500 TOPS + 20–60 TFLOPS FP16
Sensors (vision)	4× RGB-D, 2× fisheye, 2× foveal RGB, event cam	9	$1,200	total
Tactile skin	Palms, fingertips, soles, forearm tiles	—	$900	tiles + harness
Audio array	8-mic beamformer + DSP	1	$80	
Actuators	12 high-torque joints + 10 minor	22	$8,500	motors + drives
Gearboxes	Harmonic/strain-wave	10	$4,000	high-load axes
Tendon hands	Differential pulleys, sensors	2	$1,000	pair
Power	2.5 kWh pack + BMS	1	$800	extra pack optional
Chassis	Frame, panels, seals	—	$1,500	anodized Al + polymer shells
Interlocks	Tool bus connectors, IDs	—	$150	per-robot kit
Dock	Auto-align DC dock	1	$250	3.3 kW
Misc & cabling	Harnesses, fasteners, adhesives	—	$600	
Assembly & test	Line labor, fixtures, QA	—	$2,000	
Subtotal (BOM)			**$22,880**	
Overheads & margin	NRE amort., SG&A, warranty	—	**×2.0–2.5**	

Subsystem	Example Parts	Qty	Est. Unit Cost (USD)	Notes
Indicative ASP			**$45k–$57k**	aligns with paper's <$65–90k v1 range

Note: Hands/arms/soles designed as FRUs; keep spare ratios: 10% hands, 5% arms, 15% soles, 10% skin tiles.

E. Extended Test Protocols

E.1 Kitchen Micro-Suite (100-cycle)

- **Setup:** randomized loadouts of plates (Ø 26 cm), mugs (Ø 8–10 cm), glasses (highball/wine), silverware tray.
- **Procedure:** 100 unload→sort→put-away cycles; every 10 cycles inject a perturbation (wet handle, occlusion, reflective glare).
- **Metrics:** success %, grasp attempts, slips, force overshoot events (target 0), cycle time, Wh/cycle.

E.2 Laundry Deformables (20-item mixed)

- **Setup:** shirts, towels, delicates; crumple seeds (3 levels).
- **Procedure:** detect→flatten→fold→stack→put-away; 5 runs.
- **Metrics:** fold F1, retries, tactile slip count, drawer placement accuracy, operator prompts.

E.3 Cable & Plug Suite

- **Cables:** 3 m, min bend radius 25 mm.
- **Tasks:** thread through 5 guides with varying curvature; insert to Type-C/A receptacles.
- **Metrics:** success ≥95%, over-bend (0 tolerated), insertion torque profile within envelope, retries.

E.4 Social Corridor Battery

- **Agents:** 6 human proxies (variable speeds), pet bot (zig-zag), stroller.
- **Trials:** 20-min flows ×3 seeds; doorway courtesy with bidirectional traffic.

- **Metrics:** proxemics ≥99%, yield latency <500 ms median, deadlocks <3 s, e-stop ≤0.8 s p99.

E.5 Push-Recovery & Surface Adhesion

- **Impulses:** 80 N lateral (±30°), 60 N forward while stepping.
- **Surfaces:** dry tile, wet tile, dusty hardwood. Friction classifier must pre-down-mode before entry.
- **Metrics:** no falls; stabilization <250 ms; slip-detected slow mode engaged 100%.

E.6 Tool Governance Drill

- **Cases:** unregistered knife, child-present flag, hot handle @ 60 °C, tool ID spoof attempt.
- **Expected:** deny/slow/confirm paths; 0 energizations without policy; black-box snapshot on spoof detection.

E.7 Privacy & Audit Exercise

- **Task:** third-party reconstructs performance from receipts & kernel traces only.
- **Pass:** reproduce outcomes (timings, forces, success) within tolerances; no raw video used.

E.8 Update Monotonic-Safety Check

- **Method:** shadow run KLC + corridor before/after model/skill update.
- **Gate:** refusal ROC, e-stop latency, force ceiling adherence must not degrade; otherwise auto-rollback & CRL listing.

F. Field Commissioning Checklists

F.1 Home

- Map coverage ≥95%; label fixtures; define no-go/private zones.
- Policy pack applied (quiet hours, roles, tool classes).
- Run “new home in a day” smoke (reduced): unload dishwasher, fold 10 items, tidy shelf, run e-stop tests.

F.2 Back-of-House

- Geofences, signage, cell markings; tote/cart SKUs scanned.

- JSA completed; tele-ops desk reachable; operator ratio targets set.
- Run restock cell acceptance (kitting 200 picks; packout 100 boxes).

G. Notation & Units

- Forces: Newtons (N) | Torques: Nm | Power: W | Energy: Wh | Speeds: m/s | Angles: deg | Temps: °C
- Confidence thresholds: action θ_action default 0.8; human-proximity safety prior boosts SSM zone.

Takeaway: These appendices turn the paper into something you can **build, test, and ship**: precise interfaces, a compact IDL with examples, ready-to-run sim scenes, a realistic BOM, and exhaustive protocols that align with the safety and governance stance throughout.